

Communication breakdown

The UK Medical Research Council seems to have alienated a sizeable minority of the researchers that it supports. Building bridges with these disaffected individuals must be a top priority for the agency's next chief executive.

This should have been a golden year for Britain's Medical Research Council (MRC). Half a century ago, its scientists changed the face of biology. And with celebrations of the 50th anniversary of discovery of the structure of DNA now reaching a crescendo, the MRC is justified in claiming its share of the credit.

For the MRC's leaders, however, the double-helix celebrations have been muted somewhat by an outburst of criticism and controversy. Last month a sharply worded report from the House of Commons Science and Technology Committee accused the council of putting too much of its money in big projects, leaving little to fund individual grants for university-based researchers (see *Nature* **422**, 461; 2003). Now, scientists at its largest research centre, the National Institute for Medical Research (NIMR) at Mill Hill in north London, are up in arms about a draft strategic document that proposes closing the existing site, and reconstituting the institute in Cambridge.

The document, which was posted for consultation on the MRC's website (www.mrc.ac.uk) on 4 April, comes from an MRC subcommittee that was asked to develop a strategy for the council's major capital investments over the next 10–15 years. It argues that the proposed relocation will be beneficial, allowing the NIMR to operate side-by-side with the jewel in the MRC's crown, its Laboratory of Molecular Biology, which should by then be installed in a new building. The subcommittee also suggests that the translation of fundamental research into clinical advances will be facilitated by housing the NIMR in close proximity to Addenbrooke's Hospital, on the same site.

These arguments may have merit, but NIMR staff fear an ulterior motive, viewing the report as an attempt to redistribute money to the university researchers whose plight was highlighted in the Science and Technology Committee's report. The draft strategy document talks of making savings on the NIMR's current running costs, and suggests that the 'new' NIMR will be rather smaller than its current incarnation at Mill Hill. The subcommittee that drafted the report,

NIMR staff note, is dominated by the chairs of the MRC's research boards, who are based at universities.

For the MRC, tension between university-based researchers and the staff in its institutes and units is nothing new. But the current controversy comes in stark contrast to the glowing reports that the MRC received ten years ago, when the government of the day released a blueprint for the organization of British science that is still in place today. Other research councils were reorganized and reconstituted, but the MRC was left alone.

So what went wrong? Not that much, say some observers. The MRC retains an international reputation for fostering scientific excellence, and many researchers still regard it as fundamentally well run. Others, however, suggest that the MRC has begun to lose its way under its current chief executive, George Radda. They complain about a growing distance between the council's leaders and its scientific constituency, and about a lack of consultation over major funding decisions.

Talk to staff at the NIMR, and these complaints are greatly magnified. They feel undermined by rumours of the closure of Mill Hill that have been circulating for some time. The institute's research in genetics, developmental biology, neuroscience, infectious diseases and structural biology was given excellent ratings in the last external review in 2000, but NIMR staff have little trust in the objectivity of forthcoming reviews of the institute's individual research divisions — which are likely to be led by the same MRC board chairs who drafted the current consultation document.

This simmering discontent will land squarely in the lap of the next chief executive of the MRC, who will replace Radda when he steps down in September. Whoever is selected to lead the MRC forward should focus on improving communication with Britain's medical-research community and building its trust. Difficult choices, such as deciding the future of the NIMR, will always have to be made — the important thing is that they are seen to have been made fairly. ■

There are two sides to biodefence

A mystery epidemic of pneumonia serves as a timely reminder that Mother Nature is the ultimate bioterrorist.

Severe acute respiratory syndrome (SARS), which was responsible for about a hundred deaths by the time *Nature* went to press, was already roaming the globe by airliner by the time medical science realized that it was dealing with something new.

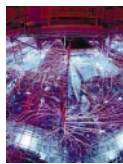
Asked at a press conference whether SARS serves as a fire drill for a bioterrorist attack, James Hughes, director of the National Center for Infectious Diseases at the US Centers for Disease Control and Prevention in Atlanta, Georgia, agreed that it does. But he pointed out that it also serves as a fire drill for another grim possibility: the next influenza pandemic.

The parallels are clear. The leading theory is that SARS is caused by a virus that jumped from animals to humans (see page 547), much like new strains of flu. And the mortality rate is similar: around 3% for SARS, compared with about 2.5% for the Spanish flu of 1918.

Thankfully, today we are better equipped to respond to the threat. The genome sequence of the prime suspect, a new strain of coronavirus, will become available any day now. This should help to reveal where the virus came from, suggest reasons for its lethality, and speed the development of rapid tests for its presence.

The genomic information would be nearly useless, however, were it not for the basic research into coronavirus biology carried out since this family of viruses was first found to infect humans in the 1960s. In the United States, vast sums of money are now being ploughed into biodefence, and nearly \$300 million of the 2003 allocation is designated for basic research, including genomics. SARS reminds us of the value of ensuring that this research is 'dual use' — relevant to fighting not just bioterrorism, but also naturally emerging diseases, which may ultimately prove more dangerous. ■





Hotting up
X-ray machine takes
small step towards
fusion power
p549



Iraq war
Coalition troops to
seek out weapons of
mass destruction
p550



Building bridges
French science
minister tries to
salvage US links
p551



Deadly waters
Chemical spill
dumps alkali in
Brazilian rivers
p552

Researchers get to grips with cause of pneumonia epidemic

Jonathan Knight, San Francisco

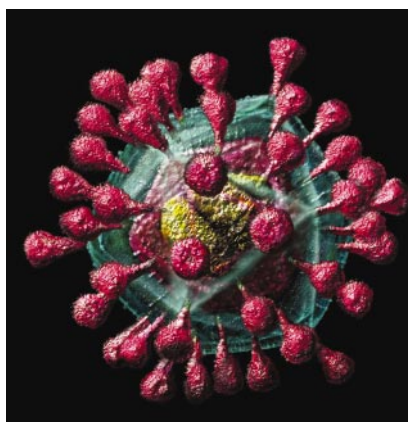
Four weeks into a major global investigation of the severe acute respiratory syndrome (SARS) outbreak, investigators around the world are still wrestling with key questions about the transmission and origins of the mystery disease.

As of 7 April, 2,601 cases of the disease had been confirmed in 17 countries, causing 98 deaths, according to the World Health Organization (WHO), which is coordinating the investigation. Epidemiologists say that the next few days will reveal the outbreak's likely extent, as health authorities discover whether people who have had no direct contact with known carriers succumb to 'tertiary' infections.

"The next week is critical," says Roy Anderson, an epidemiologist at Imperial College London who specializes in modelling the transmission of infectious diseases. "The key will be whether we start to see cases of tertiary infection."

Possible tertiary infections have already been spotted in Hong Kong. But according to David Heymann, executive director for communicable diseases with the WHO in Geneva, these could still be directly connected to the early, known cases.

In the meantime, the American Association for Cancer Research cancelled its annual conference — one of North America's largest life-sciences meetings — which was due to take place in Toronto this week, in response to concerns about the potential spread of



Prime suspect: modified spike proteins on a coronavirus may be at the root of the epidemic.

SARS between hospital staff (see below).

The investigation of SARS is continuing to focus on the role of a new strain of coronavirus — a family of viruses named for the crown of spikes they carry on their surfaces — but researchers are also finding a paramyxovirus in patients, leaving some uncertainty about the cause of the outbreak.

The coronavirus caught the attention of investigators when they saw that samples of it rapidly infected cells in culture dishes, something that other human coronaviruses do not do. Virus rinsed from the lung tissue of patients in Toronto, for example, readily infected monkey kidney cells in tests performed in mid-March. The cells began

churning out more virus and six days later began to look abnormal. "The known human coronaviruses do not infect that cell line," says Kathryn Holmes, a virologist at the University of Colorado Health Sciences Center in Denver.

Although known human coronaviruses cause nothing worse than the common cold, some that infect other mammals and birds are much nastier. Avian infectious bronchitis virus, for example, causes a severe respiratory disease and is a major problem in the poultry industry. Close contact between poultry and humans in the Chinese province of Guangdong, where SARS is thought to have originated (see overleaf), has fuelled speculation that the virus that causes it may have jumped species from poultry to humans.

Laboratory tests also suggest that such a species jump has occurred. When the coronavirus was first spotted, for example, the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, sent tissue samples and virus cultures from SARS patients to Joe DeRisi, a molecular biologist at the University of California, San Francisco, who had just finished developing a virus-detector chip. This is a glass slide on which are spotted DNA fragments from every virus whose genome has been sequenced — about 1,000 viruses — represented as some 12,000 spots.

On DeRisi's chip, any genetic material from viruses in a sample will bind to the spot containing matching genetic sequences. Fluorescent tags attached to the viral samples

Cancer conference scrapped as outbreak intensifies

Last Wednesday morning, the American Association for Cancer Research was all set for the highlight of its year — the gathering of 12,000 researchers for its annual meeting, which was due to start on 5 April in Toronto. "It was all systems go," says Marge Foti, the association's chief executive.

But late that afternoon, the meeting was cancelled — a victim of mounting concern about the potential spread of severe acute respiratory syndrome (SARS), especially among hospital staff. Toronto has been hit harder by the epidemic

than anywhere outside Asia — there have been nine deaths there — and organizers decided that holding the meeting there wasn't worth the risk.

The writing was on the wall for the meeting when the Memorial Sloan-Kettering Cancer Center in New York issued a statement advising its staff not to attend. "We told them if they did go they would not be allowed back to work for ten days after the meeting," says Chris Hickey, a spokeswoman for the centre.

Other leading US academic medical centres followed suit and this, together with an increasing

flow of cancellations from individual attendees, spelt the end for the meeting. The association will attempt to reschedule it by September, says Foti, but it is unclear how much of the meeting's US\$7-million cost can be salvaged. "It will be devastating financially," Foti says.

Other meetings were proceeding as planned last week — the American College of Physicians, for example, drew 6,500 attendees to its annual meeting in San Diego. "The benefit to practitioners of hearing the latest results on SARS far outweighs the risk," one meeting organizer says. **Tom Clarke**

R. KIGHTLY/SPL

then light up the matching spots, providing a profile of the viruses in the sample.

In the samples sent by the CDC, coronaviruses from several species lit up. The brightest spots were from avian infectious bronchitis virus and a bovine coronavirus. "This suggests that there might have been mixing between viral strains to produce this virus," DeRisi says.

The other current candidate virus in SARS, a type of paramyxovirus known as human metapneumovirus, did not light up. But because the samples were first cultured in monkey cells, it is possible that the metapneumovirus was present in the patients but didn't grow well in cell culture.

Questions about the nature of the new coronavirus will be answered within days, when virologists get hold of the complete genetic sequence now being generated at the CDC. Coronaviruses are unusually prone to incorporating material from their hosts into their genomes, so their sequences carry lots of information about their recent evolutionary past, says Holmes. It will be clear from the sequence whether, for example, bird and cow viruses combined to produce the new strain.

The sequence will also help to determine why the new virus is so much more harmful than other human coronaviruses. One possibility is a difference in the spike proteins that protrude from the cell surface, virologists say. These bind to the cell-surface receptors that pull the viral particle into the cell, so their structure determines what kinds of cell the virus infects.

Although human coronaviruses can survive on a dry surface for up to 3 hours, they are not normally transmitted through the air unless carried in the tiny droplets produced by a sneeze or cough. So close contact with a patient is the normal means of transmission, and this appears to be



A flight from Tokyo is quarantined at San José airport after three people showed symptoms of SARS.

how the current outbreak is spreading.

Physicians in Toronto and Hong Kong have used broad-spectrum antiviral agents to treat some of the more severe cases of SARS, but are not sure whether the recoveries are the result of the treatment.

And despite the intense focus on the coronavirus, researchers are aware that it may yet prove not to cause the disease, at least on its own. John Tam, director of virology at the Prince of Wales hospital in Hong Kong, has found human metapneumovirus in 25 of 53 SARS patients there, and thinks that it probably plays a role in the illness. Laboratories in Canada and Germany have also found the metapneumovirus in samples.

Metapneumovirus was first discovered in Dutch children with chest infections two years ago (B. G. van den Hoogen *et al. Nature Med.* 7, 719–724; 2001) and has since been found to exacerbate the health effects of other infections. Tam is now sequencing the virus to see if it is the same as the original Dutch isolate, but his lab has limited resources and he does not

know whether this will take weeks or months.

Much of the uncertainty surrounding the contributions of the two viruses could be resolved with the aid of direct tests for their presence in patients, to replace the indirect tests for antibodies to the new coronavirus that have been used so far. CDC officials say that they hope to have such a test for the coronavirus ready to send to US state health departments this week.

If the coronavirus is shown through animal experiments to be the cause of SARS, a direct test could prove critical to stopping its spread, says Michael Lai, who studies coronaviruses at the University of Southern California in Los Angeles.

Patients suspected of having SARS, such as passengers on a flight from Tokyo who were quarantined for two hours at Mineta San José International Airport in California on 1 April, could then be given a more definitive test. "A test will be the most important step at the current time," Lai says.

Additional reporting by David Adam in London and David Cyranoski in Tokyo.

Chinese authorities admit mishandling of Guangdong crisis

China has apologized for the way in which it handled the outbreak of the mystery pneumonia known as severe acute respiratory syndrome (SARS). The country was lax in informing its citizens and the wider world of the spread of the disease, admitted Li Liming, director of the Chinese Center for Disease Control and Prevention (CDC) in Beijing.

The apology came at a press conference in Beijing on 4 April, as international criticism mounted of China's handling of the outbreak. Liming said he hoped that the apology would reassure visitors to China. But visitors' concerns continue to mount: last week, for example, the United States began to make arrangements to bring home non-essential embassy staff.

As scientists around the world rush to identify the cause of SARS, researchers in China — who could have had a five-month head start — have

struggled to get their investigation under way.

"We have done a lot of research, but we need help in terms of materials and techniques," says one doctor in Guangdong. "We need expertise, we need primers to help us identify the disease."

The primers — short pieces of genomic sequence used to identify organisms — and other equipment should soon be shipped from laboratories abroad. But until these can be used to test samples from China's patients, it won't even be confirmed that the outbreak in China is the same as that elsewhere, as most experts now assume.

The Chinese CDC says that the coronavirus — regarded by many outside researchers as the most likely cause of the outbreak — is present in only some of the 30 patients tested during its investigation. The bacterium *Chlamydia*, however, has shown up in every case, they say. As of last

weekend, World Health Organization (WHO) investigators had yet to see any samples from a Chinese patient.

The Chinese CDC investigation has been hindered by bureaucratic divisions between the provincial government and Beijing. Even the movement of samples from the province to Beijing research institutes has been slowed down by paperwork, WHO officials say.

Although concern remains, there are reports from doctors in Guangdong that the spread of the disease is slowing. One doctor treating victims of the epidemic in the province ascribes this to successful efforts to isolate patients, improve ventilation in hospitals, and to encourage sanitary measures such as washing hands. But he also says that the disease seems to become less infectious as it progresses to secondary and tertiary infections. David Cyranoski

Infrared observatory prepares for lift-off

Tony Reichhardt, Washington

The last of NASA's 'great observatories' — a set of space telescopes conceived by the agency almost a quarter-century ago — is finally ready for launch.

If all goes to plan, the Space Infrared Telescope Facility (SIRTF) will go into orbit on a Delta rocket from Cape Canaveral, Florida, on 18 April — having survived years of delays and doubts that, its designers say, have ultimately resulted in a more capable observatory.

It is the fourth in a series of powerful NASA space telescopes covering different parts of the electromagnetic spectrum, after the Hubble Space Telescope, the Compton Gamma Ray Observatory and the Chandra X-ray Observatory.

SIRTF's infrared detectors are especially suited to studying cool objects, such as planets in the act of formation and brown dwarfs — bodies that lack the mass to ignite nuclear fusion and become stars. It will detect radiation that passes through dust that blocks observations at other wavelengths, as well as signals from very distant, fast-receding galaxies whose visible light is Doppler-shifted into the infrared spectrum.

The space observatory's 85-centimetre mirror is small by ground-based standards. But because of the sensitivity of its detectors and their placement in the cold vacuum of space, infrared observations that would require exposures of several hours from the ground "will happen in seconds", says Michael Werner, the SIRTF project scientist at NASA's Jet Propulsion Laboratory in Pasadena, California.

For Werner and others, it has been a long, tortuous trip to the launch pad. He began working on SIRTF in 1977. When NASA asked astronomers to propose instruments for the observatory six years later, the plan was to carry it into orbit inside a space shuttle. By the mid-1990s, though, the estimated cost of SIRTF — which had by then been reconfigured as a separately launched satellite — had ballooned to \$2.2 billion, far more than NASA could afford. The observatory was scaled back in size, and its observing range narrowed to wavelengths of 3–180 μm . It ended up costing \$670 million to build, and will cost a further \$500 million to launch and operate.

In some ways, the design emerged stronger. The observatory needs to be cooled by liquid helium to stop it emitting infrared noise that will interfere with its observations. Mission planners found a means of cutting the amount of helium needed by a factor of ten — instead of following a conventional orbit, the spacecraft will drift behind Earth as it circles the Sun, where the environment is cooler. That idea, plus modifications to the cooling system, should keep SIRTF running for five years or more. The previous best orbiting infrared telescope, Europe's Infrared Space Observatory, operated for only 28 months before its helium ran out in 1998.

While SIRTF's launch date kept slipping in the 1990s, infrared detectors kept improving in sensitivity. And because the spacecraft itself will be so cold, and therefore low in infrared emissions, the new detectors will be able to exploit their sensitivity to the full, says



The last of NASA's great observatories, SIRTF will provide rapid data on cool objects in the Universe.

JPL/NASA

George Rieke of the University of Arizona at Tucson, principal investigator for one of the observatory's three instruments.

Like the other three great observatories, SIRTF will be used by large numbers of astronomers around the world. It is expected to view some 20,000 targets annually, thanks to a faster pace of observation than other space telescopes: whereas Hubble looks at about ten objects a day and Chandra just three, SIRTF should rack up 50 or 60.

But for many SIRTF scientists, getting to the launch pad is a triumph in itself. Rieke recalled at a recent press conference that his team had applied to participate in the project on Friday 13 May 1983. "Somehow," he says, "we have overcome the bad aura."

♦ <http://sirtf.caltech.edu>

High-voltage shock sparks fusion from X-ray crush

Geoff Brumfiel, Philadelphia

A small burst of nuclear fusion has been induced for the first time using X-rays generated by a huge burst of electricity, according to US researchers.

Scientists from Sandia National Laboratory in New Mexico told the American Physical Society's meeting in Philadelphia on 5 April that the results were generated by the lab's Z-machine in late March. "This is the first observation of fusion from an electrical source," says Sandia physicist Ray Leeper.

Fusion occurs when two small nuclei — usually of hydrogen or one of its isotopes, deuterium or tritium — fuse together to form a larger one. Physicists try to reproduce the phenomenon, which powers the stars, using one of two techniques: magnetic confinement, which holds a plasma in place magnetically and heats it; or

inertial confinement, which uses X-rays to crush a small target so rapidly that fusion takes place.

The Z-machine attempts the latter, using energetic X-rays to crush a small sample of deuterium to form helium. The X-rays are generated by releasing a massive electrical power surge through a small cylinder of 360 tungsten filaments, which surrounds a 2-mm pellet filled with the deuterium. At the flip of a switch, scientists release a 40-terawatt pulse just 75 nanoseconds long through the filaments, vaporizing the wires instantly, and emitting X-rays that crush the pellet.

The Z-machine has been running since September 1996, but last month's experiment confirmed that emissions of neutrons — a by-product of the fusion reaction — had been detected.

The result could strengthen the case for a larger Z-machine — which Sandia has

wanted to build for years, at a cost of several hundred million dollars. It may also help with research at the National Ignition Facility, an experiment under construction at the Lawrence Livermore National Laboratory in California, which will use the world's most powerful laser to achieve inertial confinement fusion.

The fusion energy generated in the experiment is estimated to be 4 millijoules — about a billionth of the electricity put in — so the technique has some way to go in terms of efficiency.

But Per Peterson, chair of the nuclear engineering department at the University of California, Berkeley, says it could one day be possible to scale up something like the Z-machine to create an affordable fusion power plant. "Pulsed power has the potential to be by far the cheapest approach" to fusion power, he says.

Allies will expose Iraq's weapons, says Blix

Declan Butler, Paris

Hans Blix, the chief weapons inspector for the United Nations (UN), last week admitted that it will be the US-led coalition invading Iraq, and not the UN, that will take the lead in tracking down and interrogating Iraqi scientists involved in the development of any chemical, biological or nuclear weapons there.

Speaking in Stockholm at a seminar on postwar Iraq, Blix said that the coalition would stand a better chance of finding weapons of mass destruction (WMD) than

his inspectors had done, because Iraqi scientists would be more cooperative once Saddam Hussein's regime had been overthrown. "Clearly, the allies are likely to have more success than we did, as people will be less fearful of speaking up as the regime disintegrates," says Ewen Buchanan, a spokesman for UNMOVIC, the United Nations Monitoring, Verification and Inspection Commission.

The coalition forces already have Iraqi weapons scientists in their sights. "Those weapons will be identified, found, along with the people who have produced them and who guard them," General Tommy Franks, head of US Central Command, told a press briefing in Doha, Qatar, late last month. A Pentagon spokesperson confirmed that "interviews with key Iraqi scientists, technicians and administrators associated with WMD programmes may be our most powerful tools for unearthing Iraq's hidden programmes".

The United States is putting together its own 'mobile inspection teams', drawn from the armed services, Department of Energy laboratories and private defence contractors, to search for weapons and weapons scientists.

Nations that are not part of the invasion force will probably want independent UN verification of any US claims that such weapons exist or that Iraq has the capacity to

produce them. But scientists are likely to be interrogated during the conflict, and before a UN presence in Iraq has been re-established.

Unfettered access to Iraqi scientists was demanded by the UN Security Council resolution on Iraq, passed last November. But UN inspectors were able to speak privately with only 14 of the 500 listed for interview.

The US push into Baghdad this week takes its troops to where biological or chemical weapons facilities, and the scientists involved in their development, are likely to be found. The scientists, however, may have less information about actual weapons stockpiles than do military personnel, and the top US priority will be to find and secure these stockpiles. As of 7 April, no discovery of weapons of mass destruction had been confirmed.

After the war, information from the weapons scientists will be critical to any understanding of how, or if, Iraq has developed weapons of mass destruction since the first Gulf War, and whether it has shared the technology with groups outside the country.

Kofi Annan, the UN secretary general, said on 24 March that only UN inspectors could determine Iraq's weapons status, and that he expected inspections to resume after the conflict.

The US hoped to track down weapons scientists as its troops moved into Baghdad this week.

Japan aims to let industry compete for grant money

Keiko Kandachi, Tokyo

Japanese academics are nervously eyeing a government plan that would open up the country's main research programmes to industrial participation for the first time.

University researchers worry that the plan, which is being considered by Japan's highest scientific decision-making body, the Council for Science and Technology Policy (CSTP), could lead to a major transfer of resources from universities to more applied research projects in industrial labs.

As the moment, industrial researchers cannot apply for Japan's largest competitive grant programme, which distributed ¥143 billion (US\$1.2 billion) last year. The 'grants-in-aid' programme, which provides about half of all the competitive research funding available in Japan, is vital to the country's basic research. But Hiroo Imura, chairman of the CSTP's grant-reform committee, says that all funding sources should be open to all researchers, regardless of affiliation.

Japan's prime minister, Junichiro Koizumi, said at the CSTP's plenary session on 28 March that the 'grants-in-aid' programme should be "fundamentally revamped", according to a subsequent

briefing by government officials. The CSTP committee will spell out details of the revamp in a report early next month.

Advocates of the plan point out that US corporations have benefited for years from government research grants. Statistics from the US National Science Foundation show that 40% of all US federal grant money for research and development went to industry in 2001, compared with only 10% in Japan.

The CSTP report is expected to argue that the change is needed to enhance the diversity of grantees and to encourage competition between academic and industrial researchers, in line with university reforms (see *Nature* 419, 875–876; 2002).

But Takahisa Onishi, an official at the education ministry who deals with university grants, says the real motive is to help Japanese companies that have been forced by economic recession to slash their own research budgets.

The universities are furious at the plan, but may be powerless to stop it. Tadamitsu Kishimoto, president of Osaka University, who serves on the CSTP committee, predicts

that the change will effectively divert money into applied work with quick results, at the expense of long-term basic research.

Industrialists, unsurprisingly, are more upbeat. Ikunoshin Kato, president of Takara Bio, a Japanese biotechnology company, who is also on the CSTP committee, says: "It's not fair that we can't apply for large grants simply because we are in the private sector."



Ape populations decimated by hunting and Ebola virus

John Whitfield, London

Chimpanzees and gorillas in West Africa are caught in a pincer movement between hunting and the Ebola virus, researchers have warned. The bushmeat trade is threatening their populations near towns, while Ebola is killing almost every animal in some remote areas.

In a paper on page 611 of this issue, ecologist Peter Walsh of Princeton University, New Jersey, reports that populations of both species have plunged by about half over the past 20 years. Walsh calls for the conservation status of each to be shifted from 'endangered' to 'critically endangered'.

The finding that apes that live near people are in decline because of hunting confirms the long-standing suspicions of conservation workers. But the extent of Ebola's reach in the ape populations has taken experts by surprise. In one remote area where there is little or no hunting, the virus has cut the population by more than 90% since 1991.

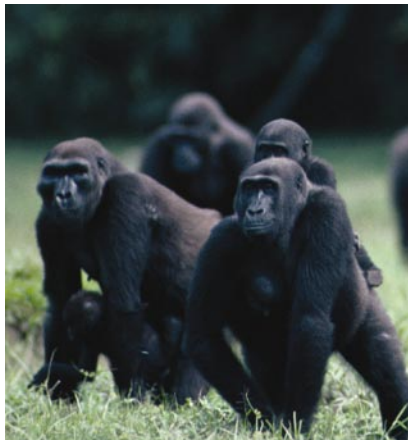
Everyone agrees that ape populations are under threat, but John Oates of the City University of New York, author of the IUCN's 1996 status survey and action plan for African apes, thinks it would be premature to reclassify them based on this evidence. It is not known whether these declines are repeated across Africa, he says, or how the picture might differ between chimpanzees and gorillas.

Government officials in Gabon, where the survey took place, say that chimpanzees and gorillas are protected there, and that every effort is made to stop illegal hunting. "The population of apes has been quite stable recently," says Pierre Ngavoura, director of water and forests at the Ministry of Water, Forests, Fisheries and the Environment.

Walsh's team counted animals by surveying overnight sleeping nests in many areas, giving a combined population for chimpanzees and gorillas. Gorilla populations in neighbouring Congo — the other remaining population stronghold of these apes in West Africa — are thought to be experiencing similarly high mortality.

The Ebola epidemic may be a result of high ape population density, says Alexander Harcourt, a primatologist at the University of California, Davis. "The normal density of gorillas is about one every two square kilometres," he says. "But in some of these regions there are ten in every square kilometre."

Another possibility, says wildlife-disease expert Andrew Cunningham of the Institute of Zoology in London, is that environmental changes, such as human encroachment on the forest, have brought apes and the virus into closer contact. "The mortality suggests



Gorillas will be missed: populations are in decline.

that there has been some trigger leading to the emergence of Ebola as an important cause of ape mortality," he says. The Ebola virus is thought to reside in an unidentified reservoir species — possibly a fruit bat or other small mammal.

Ebola spreads from apes to humans when a hunter kills and eats an animal, or when someone comes into contact with an infected ape corpse. The government in Gabon is seeking to educate its people about the risks of ape hunting, Ngavoura says.

An Ebola outbreak in Gabon killed 50 people between December 2001 and March 2002, mostly in the remote areas where the ape disease is worst. A current outbreak in Congo has killed 120, according to the World Health Organization.

A further potential hazard is created by commercial poachers who hunt bushmeat for sale in urban Africa, says Richard Rugiero, African programme officer for the US Fish and Wildlife Service. "It's a very dangerous situation in terms of global health," he says. "I don't think there's a city in the world with a West African immigrant population that doesn't receive ape meat."

Options to control the disease include cutting or digging barriers to quarantine infected populations, moving apes away from where the epidemic is raging, or culling the reservoir species, if it can be identified.

An experimental vaccine has shown good results in monkeys (see *Nature* 408, 605–609; 2000), and may be ready for human trials in a year or two, says Gary Nabel, director of the US National Institutes of Health's Vaccine Research Centre in Bethesda, Maryland. The vaccine might work for apes too, he says. "We'd very much like to use it to help."

♦ www.ApeEbolaCrisis.org

♦ www.westerngorilla.org

French minister breaks diplomatic ice with visit to US

Geoff Brumfiel, Washington

It is now ten weeks since US and French presidents George Bush and Jacques Chirac last spoke on the telephone. But as a deep freeze descends on relations between the two former allies, France's science minister, fearless former astronaut Claudie Haigneré, ventured into Washington last week to discuss scientific ties with her US counterparts.

France and the United States are split over the war in Iraq, but Haigneré said on 3 April: "In research and development, we share common objectives." In a two-day trip, she met with top officials from the National Science Foundation, NASA and the Department of Energy, and met John Marburger, President Bush's science adviser, and some members of Congress.

"There was a desire to have a constructive dialogue," Haigneré said of her visit, in which topics discussed ranged from exchange schemes for researchers to international collaboration.

Haigneré met with NASA administrator Sean O'Keefe and discussed the future of the International Space Station. She was told that he hopes that shuttle flights will be resumed by the end of the year. Issues raised at Haigneré's Department of Energy meeting included collaboration on research into alternative energy sources and the construction prospects of ITER, the international magnetic-fusion experiment, which the United States plans to rejoin (see *Nature* 421, 563; 2003).

She also met with three members of the House of Representatives: Dave Weldon (Republican, Florida), Ralph Hall (Democrat, Texas) and Bart Gordon (Democrat, Tennessee).

"In my meeting with the congressmen, the issue of Iraq was raised," she admitted, although she declined to reveal what was said.



Claudie Haigneré keeps up dialogue.

M. METZEL/AP

Spanish oil company has *Prestige* clean-up operation in the bag

Madrid Giant underwater bags could be used to remove oil from the sunken tanker *Prestige*. The move, one of three proposed plans to retrieve the oil, was announced by the Spanish government on 4 April.

Repsol YPF, the petroleum company contracted to remove the oil, plans to create a valve between the ship and the bag, and says that the oil should then rise into the bag, as it is less dense than water. If this fails, the company intends to put a canopy over the ship and catch the oil rising from it. As a last resort, the oil would be pumped from the ship to the surface. Work could begin this summer.

The *Prestige* is lying in about 3,500 metres of water, and still contains about half of her cargo of 77,000 tonnes of fuel oil. A submarine has patched up some of the holes in the wreck, but oil is continuing to leak out. No budget for the oil recovery has been set because there is no precedent for this kind of operation, Spain's deputy prime minister, Mariano Rajoy, told a news conference last week.

Spread of bird flu triggers Dutch Easter egg hunt

The Hague Dutch soldiers have been drafted in to hunt for chicken eggs this Easter in an attempt to stop the spread of a highly virulent type of chicken flu that is moving through the Netherlands.

The virus responsible for the outbreak was identified at the end of February, and control measures were put into place (see *Nature* 422, 247; 2003). These were initially successful, but last week the flu spread from the centre of the country to areas close to the German and Belgian borders in the south, crossing two rivers that were thought to be natural barriers. The transport of eggs and chickens has now been banned, and soldiers patrolling the affected area have been issued

with stop-and-search orders.

By 5 April, nearly 800 poultry farms had been infected and almost 11 million chickens culled. The virus has infected 66 humans who have been in contact either with infected chickens or with other infected humans. Although deadly to chickens, infection in humans usually causes conjunctivitis, which is sometimes accompanied by flu-like symptoms.

Life's work adds up to prize for maths master

Oslo Jean-Pierre Serre, a mathematician whose work contributed to the proof of Fermat's last theorem, has won the first Abel Prize. The £760,000 (US\$800,000) award from the Norwegian government was set up to provide a mathematics equivalent of the Nobel Prizes (see *Nature* 413, 100; 2001).

Serre received the prize on 3 April for his lifelong contributions to mathematics. Now 76, he began his academic career by studying topology, which deals with the properties of objects when they are deformed, such as by folding or twisting. In 1954, he became the youngest-ever winner of the Fields Medal, the most prestigious prize in mathematics. Later, Serre studied polynomial equations. His work contributed to the 1995 proof of Fermat's last theorem, which states that the equation $x^n + y^n = z^n$ has no non-zero integer solution when n is greater than 2.

♦ www.abelprisen.no

Chemical catastrophe poisons Brazilian rivers

Rio de Janeiro A major chemical spill has cut off the water supply to half a million citizens around Rio de Janeiro.

Caustic soda began leaking from a wood-pulping factory on 28 March, according to the environmental group WWF-Brazil. More than 1.2 billion litres had emptied into two rivers — the Paraibo do Sul and Pomba — before the leak was shut off three days later. Caustic soda, or sodium hydroxide, is used in paper mills at concentrations of



Dead in the water: caustic soda from a wood-pulping factory is killing fish in the Pomba river.

50% or higher, according to Mary DeVany, president of DeVany Industrial Consultants in Vancouver, Washington. "At 50% concentration, it has a pH of 14. A 10% solution will eat the skin right off your body," she says.

Locals have been told not to drink or bathe in water from the rivers, and the rise in pH is killing plants and animals living in or near the affected water. WWF-Brazil says the incident shows that the government is unable to deal with such accidents, as the spill would have been spotted by the water-monitoring systems in use in other countries.

Science minister's gift triggers political storm

London David Sainsbury, the supermarket magnate turned British science minister, is at the centre of a political row after donating £2.5 million (US\$3.9 million) to the ruling Labour Party in whose government he serves. The gift — the largest ever made to the party — takes Sainsbury's donations to Labour to over £8.5 million since 1999.

The Labour Party revealed on 31 March that it had accepted the money. Critics inside and outside the party say that it should be sent back. "The fact that Lord Sainsbury is a government minister raises real questions over this donation," says Theresa May, who chairs the opposition Conservative Party.

Both the government and Sainsbury — whose family's wealth is believed to run to billions of pounds — insist that the donation is unrelated to his position. Sainsbury regularly gives money to other causes and, through the Gatsby Charitable Foundation, his family supports work in several areas, including the developing world and mental health.

Seeing the big picture

Berlin The dramatic architecture of the Sony Centre in Berlin's Potsdamer Platz is the new backdrop for the Artwork Earth exhibition, an imposing series of oversized photographs of the Earth seen from space.

The travelling exhibition — curated by the Helmholtz Society, which organizes Germany's 16 national research centres — consists of a series of 3.6-metre-high images that are floodlit at night, and will be seen by 60,000 people each day. It opened last year to mark Germany's Year of the Geosciences, and arrived in Berlin on 3 April.

Null and void

Many scientific studies produce negative results that never see the light of day. Is progress in some disciplines being hampered by researchers' tendencies to consign these data to the bin? Jonathan Knight investigates.

Although scientists clamour to publish the results of successful experiments, they are less excited about trumpeting those that simply confirm the null hypothesis — that a particular genetic marker isn't associated with an inherited disease, for instance, or that there is no difference between mice given a candidate drug and those in the control group.

Whether a result of an eagerness to move on — or perhaps, in some instances, a desire not to reveal to competitors the avenues they have been fruitlessly exploring — most researchers don't bother to write up negative results. Even when they do, journals might be unreceptive. Unless a paper convincingly overthrows a widely held belief, negative findings tend to be of less interest than positive ones.

But what is the cost to science of all these data languishing in the bin? How many postdoc years and scarce grant funds are wasted on projects that have failed previously in other labs? And is our scientific understanding in some cases biased by a literature that might be inherently more likely to publish a single erroneous positive finding than dozens of failed attempts to achieve the same result?

Answering these questions is extremely difficult. In some fields, awareness of negative results tends to spread rapidly by word of mouth, even if the data are never published. If a cell line fails to behave as described in a high-profile paper, for instance, the news tends to spread among the biologists who need to know, whether or not anyone actually publishes a paper refuting the original discovery.

Nevertheless, many researchers interviewed for this article suspect that their disciplines would benefit if negative results were to get a public airing. At present, the obstacles to disseminating negative findings make it difficult even to assess the extent of the problem. "It is hard to see what the bottom of the iceberg is like when you are sitting on top of the water," observes Helen Colhoun, a genetic epidemiologist at University College London.

Awareness of the problem is growing.

Over the past few years, a handful of journals and online repositories dedicated to negative results have been launched — with varying degrees of success. In certain fields, some scientists are even arguing that a requirement to reveal negative results should be made a condition of publishing a positive finding.

Gold standard

In Colhoun's discipline, the problem is particularly acute. The postgenomic era has seen an explosion of 'gene association' studies, in which researchers screen large numbers of people for thousands of genetic markers. Their aim is to see whether some of the markers seem to be inherited alongside a disease, which is taken as evidence that a gene conferring susceptibility to the condition lies nearby. But just as gold prospectors keep the nuggets and throw the

pebbles back into the stream, those engaged in the new genetic gold-rush tend to report only positive associations, leaving the rest to be panned through by others again.

Worse, it turns out that many of the positive results that have been published may be errors. Last year, a team led by geneticist Joel Hirschhorn of the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, reviewed the literature on 166 common genetic variants that had been linked to diseases such as heart disease or acne at least once, and which had been subject to association analysis at least three times. He found that there were consistent results for only six of the variants¹. This suggests that false positives and false negatives are all too easy to come by — and because there tends to be a bias towards publishing positive associations, it stands to reason that many genetic links to disease described in the literature are wrong.

The practice of shelving negative results also leads to problems in other fields. Two years ago, Britain's Animal Procedures Committee, which advises the UK government on its policies on research involving animals, raised the concern that scientists may be duplicating experiments that had already failed in other labs. The committee recommended that the Home Office, the government department responsible for issuing licences for animal research, require its licensees to share their negative results, possibly through a government website. But officials argued that publishing research findings is not the government's job, and the issue remains unresolved.

Non-publication of negative results may also be skewing the debate over the safety of transgenic crops. Trials of genetically modified plants overwhelmingly reveal no adverse environmental consequences or health effects, argues Alan McHughen, a plant biotechnologist at the University of California, Riverside, and the results generally go unpublished. "Journal editors say: 'So what?'" McHughen says. Not that he blames them — he wouldn't



GETTY IMAGES



Helen Colhoun and Scott Kern believe researchers would benefit from greater access to negative results.

want to pick up *Nature* and read a procession of negative, unsurprising findings. But as the trials tend to be catalogued in obscure government documents, the public and scientists outside the field are often unaware of them.

Can anything be done? Clearly, journals that seek to maximize their visibility will continue to publish only high-impact papers. Occasionally a negative result falls into that category. On 27 February, *Nature* published a physics paper that ruled out certain types of string theory by searching for deviations from Newton's inverse-square law and finding none². And *The Lancet* recently published a large study that failed to confirm a previous hypothesis that certain versions of the gene for apolipoprotein E make smokers more susceptible to heart disease³.

Positive steps

But these are exceptions. To handle the steady stream of lower-profile negative findings, some scientists are setting up their own publishing efforts — with mixed results. Bjorn Olsen, a cell biologist at Harvard Medical School, for instance, established the *Journal of Negative Results in Biomedicine* (*JNRBM*) last year. The main requirement is that the results be reproducible. Beyond that, anything biomedical goes, from failed clinical trials to reagents that don't work as advertised.

Submissions to the *JNRBM*, which is published online by London-based BioMed Central, go out for peer review only if they are deemed interesting by the journal's editorial board. This should prevent the *JNRBM* from becoming a laundry list of experiments with predictably negative outcomes, says Olsen. Since the journal went live in November, it has received 11 submissions, 7 of which have gone out for review. Three of these have been

accepted and two are now online, Olsen says.

A modest beginning, perhaps, but better than a similar effort in computer science, set up in 1997. In an article⁴ in the *Journal of Universal Computer Science*, editorial board member Lutz Prechelt of the University of Karlsruhe in Germany announced a new section of the journal to be called the Forum for Negative Results (FNR). He argued that valuable insights in computer engineering are lost when people discard their failed solutions to problems, rather than reporting them. But since then, there has not been a single submission, leaving Prechelt to suspect that computer scientists just don't like facing their failures. "Maybe I should write up a submission to FNR describing the concept as a negative result," he jokes.

Another publication that encourages the submission of negative findings is a new biotechnology journal. The Paris-based International Society for Biosafety Research, of which McHughen is a board member, recently launched *Environmental Biosafety Research* as a forum for publishing field trials of transgenic crops, including the majority that show nothing alarming or surprising. The first issue, which appeared in October, included four research articles, one of which reported the negative result that an insecticide-producing maize did not harm non-target species.

Although journals may be part of the

No one wants to pick up a journal and read a procession of negative, unsurprising results.

answer, they won't work well for fields such as gene association, in which negative results outweigh positives by orders of magnitude. "No one can read 150 papers and remember what they read," says Colhoun.

In such cases, presentation of negative findings in a more abbreviated form on the Internet seems the obvious answer. To this end, Colhoun has assembled a group of colleagues to discuss possible approaches, with a view to publishing the results of their deliberations in *The Lancet*, which has taken an interest in the issue. One important issue is ensuring that sufficient experimental details are provided to allow negative results to be interpreted. Colhoun's group is, for instance, considering recommending something akin to the MIAME (minimum information about a microarray experiment) standards established last year to aid the comparison of gene-expression studies using DNA chips.

Still, the availability of an online database into which scientists can deposit their negative results does not guarantee that it will be used, as Scott Kern has found. A cancer researcher at Johns Hopkins University School of Medicine in Baltimore, Maryland, Kern set up *NOGO*, which stands for the *Journal of Negative Observations in Genetic Oncology*, on his website six years ago. Although styled as a journal, *NOGO* is a repository for brief reports of negative results, including those that have been published elsewhere. When he set up the site, Kern provided a simple form for submitting negative results about genes suspected to be involved in cancer, approached colleagues at meetings and distributed flyers. Reactions were very positive, but contributions never rose above a trickle.

So one of the tasks Colhoun has set for her working group is to come up with a system of carrots and sticks to prise out negative data from the genetic-epidemiology community. One idea might be for journals publishing positive findings to require authors to make any negative data available as a condition of publishing a positive gene association, says Mark McCarthy, one of Colhoun's colleagues at University College London, and a member of her working group.

Whether journal editors will buy into that idea is unclear. But McCarthy is optimistic that a cultural shift is under way. "People are starting to realize the benefit of looking at other people's negative data," he says. ■

Jonathan Knight writes for *Nature* from San Francisco.

1. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. *Genet. Med.* **4**, 45–61 (2002).

2. Long, J. C. et al. *Nature* **421**, 922–925 (2003).

3. Keavney, B. et al. *Lancet* **361**, 396–398 (2003).

4. Prechelt, L. *J. Univ. Comp. Sci.* **3**, 1074–1083 (1997).

Journal of Negative Results in Biomedicine

► www.jnrbm.com/start.asp

Journal of Negative Observations in Genetic Oncology

► www.path.jhu.edu/NOGO

Environmental Biosafety Research

► www.edpsciences.org/ebr

The light fantastic

A decade ago, holographic systems promised to revolutionize data storage. The early hype may have evaporated, but the technology quietly progressed, and working devices are now on the market. Mark Haw reports.

Science is suffering from information overload. Astronomers struggle with huge image files from their telescopes; high-energy physicists are bombarded with numbers; and a new kind of scientist — the bioinformatician — has emerged over the past decade to help to make sense of vast biological databases. Some companies can now turn a profit simply by storing other peoples' data¹.

Computer memory is cheap, so this data glut is of little concern to most researchers. Yet the way in which we store information literally only scratches the surface of what is possible. Humans have been recording data on surfaces since the era of clay-tablet accounting systems. The printing press, photography, magnetic tape and compact discs (CDs) have each transformed data storage. But if we want to go further, and cram more data into less space, we may have to move beyond surface technologies, and actually store information inside materials — in the form of a hologram.

Ten years ago, with magnetic-disk technology seemingly reaching its maximum potential, holographic data storage seemed destined for the centre stage. But as the capacity of magnetic disks defied expectations, interest in holographic alternatives waned, and big companies ceased their research efforts. A few smaller firms, however, as well as some academic researchers, kept the faith. Spurred on partly by the advent of digital videos and problems with archiving

feature films, commercial holographic storage options are beginning to emerge.

The potential advantages of holograms are indisputable. Opening up the inside of a material for storage, rather than just using its surface, yields huge improvements in capacity. For example, InPhase, a holographic-technology company based in Longmont, Colorado, is marketing a disk that is similar in size to a standard CD but can store 200 gigabytes — a 200-fold improvement on optical CD technology. Data can also be read more quickly from such systems. And because of the way in which information is stored, useful new types of searching technique are possible.

Bright idea

The basic technique for such storage was proposed in 1963 by Pieter van Heerden, who was working on holographic technologies at Polaroid's labs in Cambridge, Massachusetts². The ones and zeros that make up the data set are first split into two-dimensional pages of data — lines of light and dark pixels displayed on a screen. Individual pages are then illuminated from behind with a laser, producing a projection of the page known as a data wave.

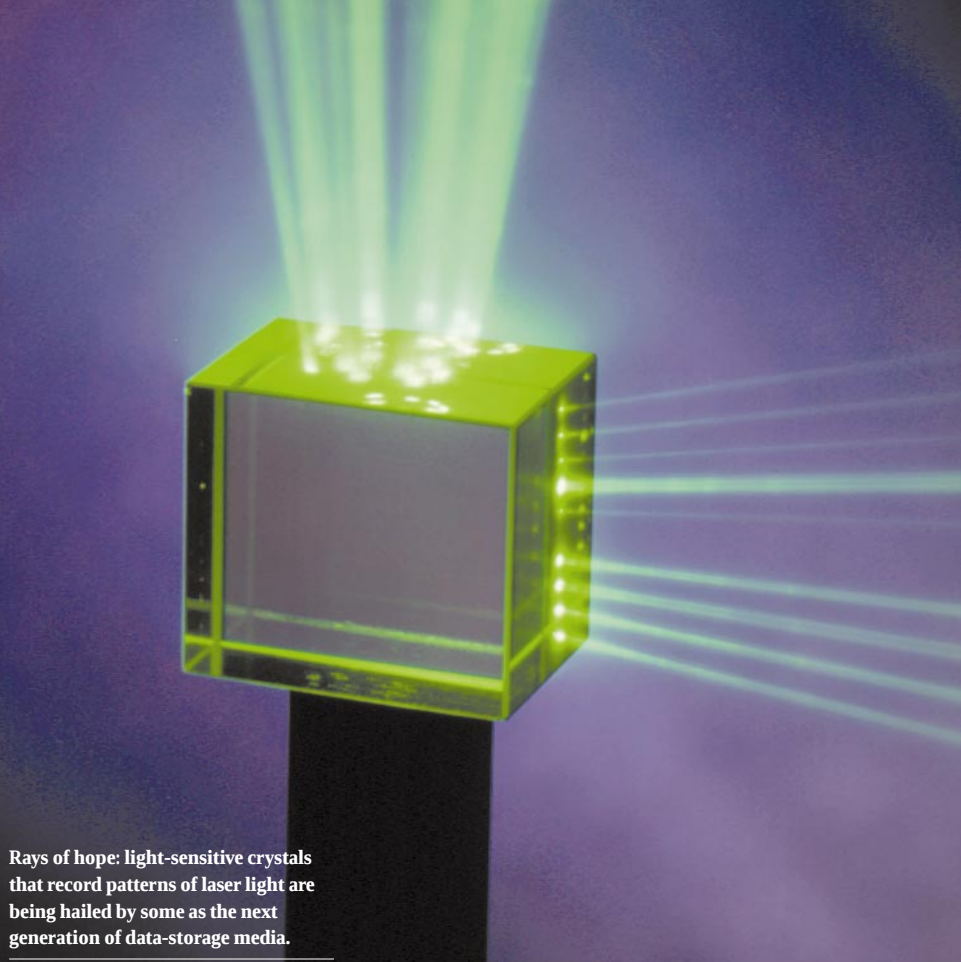
To store a page, the data wave is projected onto a block of transparent material, together with a second pattern of dark and bright pixels called the reference wave. The peaks and troughs of the two waves interfere to form a new pattern and, if the block is

made from a substance that reacts to light and dark in different ways, the interference pattern is stored inside the material.

Taking a single page and turning it into a three-dimensional pattern isn't much use in itself. The trick of holography is that, by using a different reference beam for each data page, many pages can be stored in the same volume. This is possible because the angle between the data and reference waves determines how the peaks and troughs of the two beams line up, and hence dictates the interference pattern. By varying this angle for different pages, an entire data set can be stored in a single block.

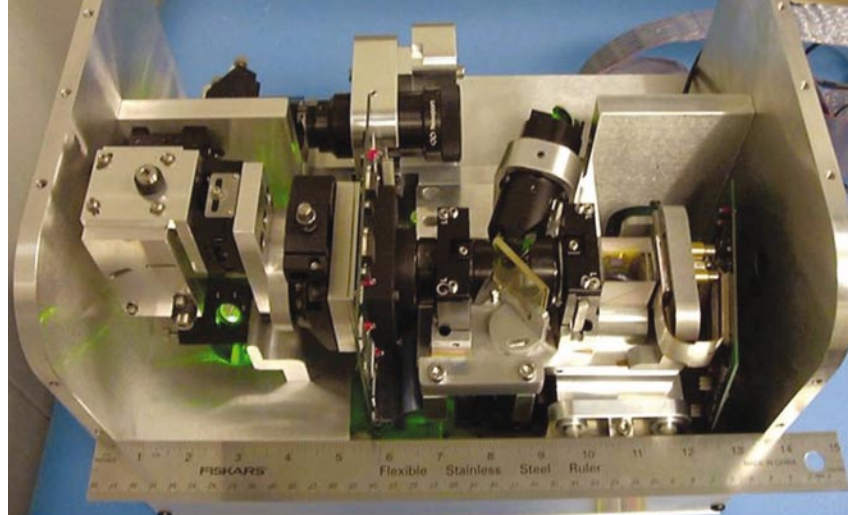
In the same way that a radio can be tuned to a single frequency amid the overlapping transmissions from other stations, a third kind of light beam — the reading wave — can be used to extract individual pages from the many patterns stored in the block. Any beam directed at the block will be scattered by the patterns, but if the reading wave has the same wavelength as a particular reference wave, and is beamed into the block at the same angle, it will pick out the data page stored using that reference wave. The scattered light will be a recreation of the data page, which can be read into a computer by projecting it onto a video camera.

Van Heerden's technique is known as holographic data storage, as the method for storing the pages is similar to that used to create holograms — the different pages of data are analogous to the different views of an



Rays of hope: light-sensitive crystals that record patterns of laser light are being hailed by some as the next generation of data-storage media.

IBM RESEARCH, ALMADEN RES. CEN.



One for posterity: holographic recording systems such as this one could be useful to movie archivers.

object that are stored in a hologram. A whole page of data is read or written at once — conventional memories read data bit by bit — so retrieval speeds are higher. And the angular tuning required to retrieve different data pages can be controlled by mirrors that can be deformed by electric signals, eliminating the need for mechanical parts — another benefit when speed is of the essence.

Sharp contrast

This straightforward technique might already be in common use, were it not for the drawbacks associated with light-sensitive materials. The basic requirement is simply that the material reacts to the light-and-dark pattern of the overlapping data and reference waves, and stores that pattern. Transparent light-sensitive polymers, or photopolymers, are one class of materials that can do the job. In these substances, the pattern causes small molecules, or monomers, in bright regions to join together into polymers, whereas dark regions remain unreacted. Monomers and polymers scatter light in different ways, so the reaction leaves a pattern that can be retrieved using the reading wave.

Photopolymers have been around for some time — they are used for the holograms on credit cards, for example. But researchers who are trying to use them in holographic storage systems have been plagued with problems that they are only just beginning to overcome. One issue is that the polymerization reaction produces only a small difference in light-scattering properties. High-powered lasers are then needed to read the pattern, making the system unsuitable for desktop applications.

Adding contrast-increasing substances to the polymer could help to improve matters. Last year, for example, optics researcher Yasuo Tomita and his colleagues at the University of Electro-Communications in Tokyo showed that adding particles of titanium dioxide to methacrylate, a Perspex-like polymer, could significantly improve contrast³. During recording, monomer levels fall in the bright areas, causing monomers to flow in from dark regions. Titanium particles aggre-

gate in spaces left behind by these monomers, and this enhances contrast between the light and dark areas. Less-powerful lasers can then be used to read the memory, although the contrast is still too low for desktop systems.

A second major problem is in producing materials that are thick enough to store lots of data. In credit-card holograms, for example, the photopolymers are held in a plastic film that typically cannot be made thicker than a fraction of a millimetre. High data density in a holographic memory requires thicker volumes, as the number of pages that can be stored depends on how many wavelengths of the reference beam will fit into the material used. Thinner pages mean fewer wavelengths and lower capacity.

One potential solution comes from the work of optics researchers Pavel Cheben, of the Institute for Microstructural Sciences in Ottawa, and Maria Calvo, based at the Universidad Complutense in Madrid. Cheben was originally looking for materials for telescope optics. After discovering that commercially available photopolymers could not be made thick enough for his requirements, he realized that other materials could be used to hold the monomers. "The idea surfaced one day," says Cheben. "Why not use silica glass, which can be fabricated to virtually any thickness?"

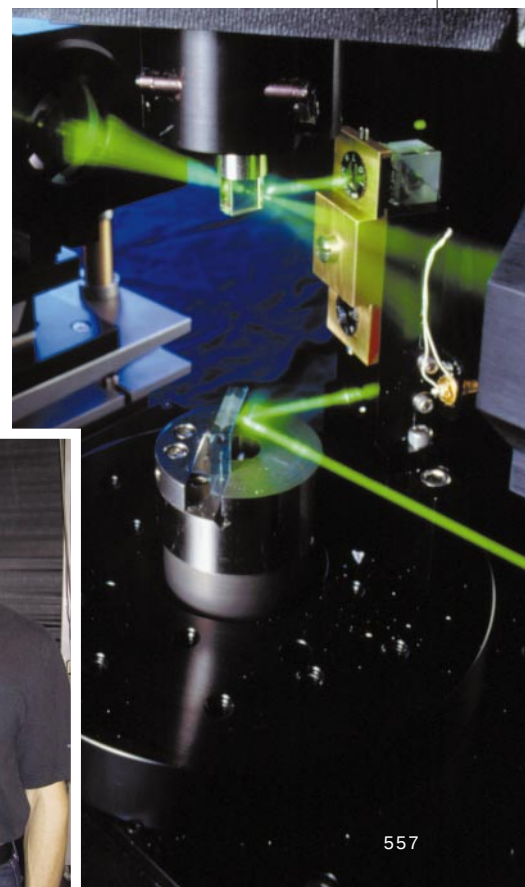
Lighting the way: 'photorefractive' materials (right), pioneered by Karsten Buse (below) for use in data storage, could give rise to robust yet rewritable media for holographic information.



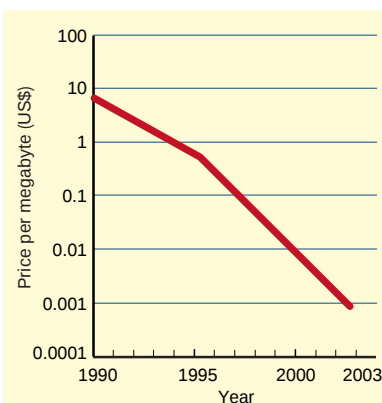
Last year, Cheben and Calvo described a holographic material contained by a porous matrix of silica glass⁴. The material has so far been used only in laboratory tests, but a 1-mm block of the substance seems to be able to store about 100 interference patterns. And like the titanium approach, the new substance increases the toughness of the material and helps to combat the third big headache involved in photopolymer memories: the disastrous warping of the material that can occur during recording and subsequent use.

Such developments show that the problems with polymer memories can be overcome. But for some researchers, there is a bigger prize worth chasing: a material that can not only store patterns, but can also erase them, and so can be used again and again. The current frontrunners are photorefractive crystals — materials in which light redistributes electrons. In these crystals, the electric field associated with the data and reference waves moves electrons from light to dark regions, rearranging charge so that it corresponds to the interference pattern.

Holograms were first stored in this way in a lithium niobate crystal at Bell Laboratories in Murray Hill, New Jersey, in the 1960s (ref. 5). But the record faded after ten minutes of illumination by a reading beam. The charge in the crystal is just as sensitive to the reading beam as it is to the recording beam, so reading the data washed out the charge pattern. Even normal room lighting erased the memory within a few hours — not much use for archiving cherished home videos.



IBM RESEARCH, ALMADEN RES. CEN.



Attractive prospect: the cost of magnetic storage has fallen rapidly in recent years.

Ingenious ways of fixing the holographic memory have since been explored, and one method looks like it might be up to the job. In 1998, Karsten Buse, an optics researcher then at the California Institute of Technology in Pasadena, described how his group added manganese atoms to a lithium niobate crystal. These act as both suppliers and traps of electrons. Because the atoms hold onto their electrons strongly, a high-energy ultraviolet beam is needed to release the electrons. The electrons move in response to the interference pattern, but some are recaptured by the manganese traps. The distribution of recaptured electrons is a copy of the pattern, and is stable because reading with a lower-energy beam does not release these electrons⁶.

Such memories are rewritable, as the ultraviolet beam can release the trapped electrons and wash out the recorded copy of the data when this is required. The idea of using a high-energy beam to switch recording on and off, while a low-energy beam allows non-destructive reading, arose at Bell Laboratories in the 1970s (ref. 7). The lasers and optical components of the time were too expensive and of insufficient quality for the technique to be feasible — but this is not the case today. Researchers are now trying to reduce the time taken to store data pages — using an affordable 1-milliwatt device, this currently takes more than ten minutes per page. Buse, now at the University of Bonn in Germany, continues to work on holographic systems, and is confident that problems such as the exposure time can be ironed out.

So are we entering the era of desktop holographic data storage? Not yet, says Hans Coufal of IBM's Almaden Research Centre in San Jose, California. "At the beginning of the 1990s people thought we were very close to the limit of magnetic-disk capacity," he says. "IBM began seriously looking into holographic data storage." But the hard-drive capacity of a standard desktop computer has doubled annually for the past five years. There have also been difficulties in making optical components cheap enough to compete with magnetic sys-

tems, whose price has consistently dropped for more than a decade (see graph, left). Coufal says this is enough to convince IBM that holographic memories are not yet ready for entry into such a fiercely competitive market.

Other researchers are more optimistic. According to Buse, the doubts are an economic rather than a technical problem — money is being channelled into tried-and-tested technologies, rather than riskier alternatives. "Industry trends are currently not in favour of holographic storage," he says. "Large funds are invested to improve the conventional technology and not to boost holography."

Screen tests

But whereas big players such as IBM have dropped out of the field, smaller companies see potential in holographic technologies. InPhase, as well as Aprilis of Maynard, Massachusetts, are marketing holographic storage systems based on photopolymers. Neither will reveal details of the polymers they use, but Aprilis claims to have reached storage capacities as high as 15 gigabytes per square centimetre, and InPhase says that Sony is interested in developing its technology. Lisa Dhar, InPhase's vice-president of media development, says that the company's 200-gigabyte disk can be read at 20 megabytes per second, compared with the 5 megabytes per second that can be obtained from CDs.

The high costs involved in manufacturing the optics for the systems, together with the fact that powerful lasers are needed for recording, mean that neither company is yet aiming at the domestic market. InPhase says that it is talking to movie-industry companies such as Universal Studios and Technicolor about long-term safeguarding of video material. "There's no acceptable long-term solution," explains Dhar. "Videotape has an archive life of about ten years." The current fix is to record video onto longer-lasting photographic film. Dhar claims that InPhase's

holographic technology will be nearly 50 times cheaper, and that 60 minutes of uncompressed video — which would require 1,600 metres of film or 100 DVDs — could be stored on just three holographic disks.

Other possible image-archiving applications include medical imaging, oil exploration and satellite imagery. But all of the companies involved admit that it is too early to say which applications will prove profitable. "It's a bit difficult to predict how all this will shake out," says Glenn Horner, vice-president of business development at Aprilis. "Much of the story has yet to be written."

What might eventually tip the balance towards holographic data storage is one feature that cannot be matched by other methods. Just as a data wave can be retrieved using a reference wave, a reference wave can be retrieved using a data wave. Known as content-addressable searching, this allows users to find out if a particular data page is present by illuminating a memory with the corresponding data wave. If the page is there, the original reference wave will be recreated. The method also allows for fuzzy searches — if there are several pages similar to the search page, various recreated reference beams will appear, with intensities that depend on how closely each matches the search page.

This presents exciting possibilities. In computer vision, for example, the method can be used to detect matches between images of the same scene viewed under different lighting conditions. In 1995, a team led by Dmitri Psaltis at the California Institute of Technology created a robot that navigated around a building using content-addressable searching to compare real-time camera pictures with images of known locations stored in its holographic memory⁸.

Work in this area continues, with some researchers linking holographic memories with artificial neural networks — computing devices based on the structure of the human brain. The combination provides rapid image recognition, and can be trained to recognize different types of object⁹.

After a long gestation, recent advances in materials, optical technology and laser availability are at last combining to make holographic memories a reality. Although you won't see one on your desktop just yet, even the most sceptical observers think that the technology will blossom in the long run. Eventually the elusive limit to surface storage capacity will be reached, after which the third dimension may be the only way to go. ■

Mark Haw is a freelance science writer and physicist at the University of Edinburgh.



Dmitri Psaltis' robot uses a searchable memory of holographic data to help navigate its way home.

1. Dennis, C. *Nature* **417**, 14 (2002).
2. van Heerden, P. J. *Appl. Optics* **2**, 387–392 (1963).
3. Suzuki, N. et al. *Appl. Phys. Lett.* **81**, 4121–4123 (2002).
4. Cheben, P. & Calvo, M. L. *Appl. Phys. Lett.* **78**, 1490–1492 (2001).
5. Chen, F. S. et al. *Appl. Phys. Lett.* **13**, 223–225 (1968).
6. Buse, K., Adibi, A. & Psaltis, D. *Nature* **393**, 665–668 (1998).
7. von der Linde, D. et al. *Appl. Phys. Lett.* **25**, 155–157 (1974).
8. Psaltis, D. & Mok, F. *Sci. Am.* **273**, 70–76 (1995).
9. Frauel, Y. et al. *Appl. Optics* **40**, 5162–5169 (2001).

Classical plant breeding is the route to food security

The right technology in uncertain conditions is one that can manage many subtle traits.

Sir — Your unusual and perceptive News Feature “Crop improvement: a dying breed” (*Nature* **421**, 568–570; 2003), on the importance of classical plant breeding, contains an error concerning intellectual property rights applied to plants. Plant breeders’ rights, called ‘plant variety protection’ in the United States, allow and defend the right of plant breeders to use protected varieties as parents in further breeding, contrary to the impression given in your feature. The patent system, not plant breeders’ rights, prohibits the use of protected varieties as parents in further breeding. Moreover, utility patents on plants became important only after public plant breeding had already declined.

That decline reflects a sea change in US public policy. From the Civil War until quite recently, agricultural research was seen as a public responsibility, creating public goods. Then, decades of sustained prosperity after the Second World War

gave rise to a belief that this was no longer necessary. It was also believed that market forces would supply all the goods that were needed, and that the public sector would be competing with the private sector.

During the same decades, the formerly rural US population became suburban and urban. Public goods of concern to today’s policy-makers are the ones that preoccupy urban citizens. These are the reasons behind the decline in support for public plant breeding (see K. J. Frey, *National Plant Breeding Study*, Iowa Agric. Home Econ. Exp. Station, Ames, Iowa; 1996), as correctly pointed out in your News Feature.

Much as we may consider agriculture to be passé — if we think about it at all — our urban lifestyle depends on a secure supply of abundant, inexpensive, safe and nutritious food. As pointed out in your News Feature, classical plant breeding is the only technology that is currently capable of delivering the secure harvests we require.

This is because only plant breeding can manage many subtle traits at once, as needed to confront challenges of climate change, newly emerging pests and other uncertainties.

Public plant breeding is also a good investment. The rate of return to investment in public plant breeding is 35% for potatoes in the northwest United States, realized in reduced cost of production and improved food quality as well as in increased yield, according to A. A. Araji and S. Love (*Am. J. Potato Res.* **79**, 411–420; 2002). These results are typical of studies in other crops.

Ann Marie Thro*, Paul Zankowski†

*Cooperative State Research, Education, and Extension Service, US Department of Agriculture, 800 9th Street SW, Washington DC 20024, USA

†Plant Variety Protection Office, Agricultural Marketing Service, US Department of Agriculture, 10301 Baltimore Boulevard, Beltsville, Maryland 20705-2351, USA

Building bridges through cooperation in science

Sir — I was heartened to see *Nature*’s Editorial “Time to unite Islam and science” (*Nature* **422**, 99; 2003). Global events since 11 September 2001 have given a new sense of urgency to the Third World Academy of Science’s efforts, over the past 20 years, to make science a core part of national policies in countries in these regions.

Although there is woefully inadequate investment in science and education in many predominantly Muslim countries, along with poor rates of economic development and a growing sense of despair and frustration, discussion should focus not on where to assign blame but on strategies for easing the tension and pointing to a better future.

Two aspects of the workshop held in Trieste in March (see *Nature* **422**, 101; 2003) that were described in your Editorial offer small but useful clues on how to move forward.

First, the workshop was jointly funded by the US National Academy of Sciences and, to a lesser extent, the Organization of Islamic Conference Standing Committee on Scientific and Technological Cooperation, and the Islamic Educational, Scientific and Cultural Organization. This is an example of North–South scientific cooperation that could have a positive impact on future relations between these countries on issues related to science and

development. More partnerships are needed between cultures that have too often chosen to emphasize their differences rather than their similarities.

Second, the workshop recommended that scientific communities in Muslim countries should increase their mutual cooperation by the creation of a network of science academies for scientific exchange, education and training, and by joint research programmes in areas of critical concern to Muslim countries. This is an example of South–South cooperation that holds the key for future success in efforts to advance science-based development.

North–South and, more important, South–South scientific cooperation could help forge alliances of mutual concern that will help to ease apprehension and distrust among nations.

Mohamed H. A. Hassan

Third World Academy of Sciences,
Strada Costiera 11, 34014 Trieste, Italy

Poetic pride in steps towards knowledge

Sir — In their fascinating Insight Review “Inflammation and cancer” (*Nature* **420**, 860–867; 2002), Lisa M. Coussens and Zena Werb traced knowledge of the relationship between inflammation and cancer back to the hypothesis proposed by the German physician Rudolf Virchow: that chronic inflammation caused by certain irritants can lead to cancer. This

was later proved correct by two Japanese scientists, Katsusaburo Yamagiwa and Koichi Ichikawa, who showed that repeated painting of coal tar onto rabbits’ ears causes carcinomas.

Yamagiwa was delighted with their results and composed a haiku: “Cancer was produced. Proudly I walk a few steps”. Their paper was originally written in German and published in a journal of the Imperial University of Tokyo in 1915. An English version was published in 1918.

In his Nobel lecture of 1966, Peyton Rous acknowledged the significance of this work. He wrote, “And then in 1918 came the epoch-making discovery by Yamagiwa and Ichikawa that tarring rabbit skin will cause tumors to arise. This opened an era of rewarding search for other chemical agents — and physical as well — which do the same. It is an era so far from done as to have been the main theme a few months ago of the International Cancer Congress held in Japan; and it was chosen with good reason since some of man’s habits, many of the occupations through which he earns his living, and the mischances taking place in his own body can prove fatal through the growths they induce unless something is done to ward them off. All of these oncogens [sic] are initiating in character, and some can be dangerous promoters too, if the exposure to them is long.”

Min-Liang Wong

Department of Veterinary Medicine, College of Veterinary Medicine, National Chung-Hsing University, Taichung 402, Taiwan

Tapping the power of small institutions

Now is the time to create a virtual, global network for health research.

Gerald T. Keusch and
Carol A. Medlin

Research and development (R&D) into health is increasingly recognized as critically important in the fight against the major causes of morbidity and mortality among poor people in developing countries. Although this may seem self-evident, it is a striking statement, as health (and health R&D) has only been recognized as a global development goal since 1993, when *World Development Report: Investing in Health*¹ was published. After this report, a World Health Organization (WHO) committee² suggested priorities and made recommendations for the mobilization of national, international and private-sector R&D support to ensure scientific and technological progress in these areas.

Most recently, the WHO Commission on Macroeconomics and Health (CMH)³, chaired by Jeffrey Sachs, made concrete proposals to push this agenda still further. Significantly, the commission defined health R&D as a “global public good”⁴, thus emphasizing the collective benefit that could be derived by all countries from such investments in knowledge for better health. It mandated “research and development into diseases (often, though not exclusively, tropical diseases) that are concentrated in poor countries”, and noted that enlightened self-interest is as compelling as humanitarianism as justification for action.

Putting theory into practice

The status of R&D investments on the international health agenda has been elevated by the designation of health research as a global public good. Now, we urgently need to make it happen. Unfortunately, therein lies the problem, as the theoretical underpinnings of such global public goods with respect to productive capacity, delivery systems and finance are relatively weak. From an economics perspective, for example, we know that it makes little difference whether the investors in R&D are many or few, public or private, or based principally in the north or the south. Fundamentally, the only thing that matters is that the investment is made.

From an ethical, institutional and practical perspective, careful consideration is essential in the organization of any global R&D strategy — what should the priorities be? How should decisions be made? Who should participate? Who should take the lead? If it is to be a truly global effort, it must emerge and take shape by global discussion and debate — yet not every nation has a robust research



Thinking big: a virtual network of health-research institutions could readily tackle global issues.

system that could be engaged for this purpose.

Some advocates have suggested creating a new global institute for health research, along the lines of international projects such as the particle accelerator at CERN, the European Laboratory for Particle Physics near Geneva. The International Vaccine Institute in Korea comes close to this concept, although it is still limited in scope and resources. The WHO's tropical disease research programme and the Multilateral Initiative on Malaria, with a rotating secretariat located at various institutions, are other 'prototypes' for global health R&D funding. Others, seeking to generate resources and avoid creating yet another UN-style bureaucracy, have proposed a global fund for health research with an independent, highly streamlined secretariat, analogous to the Global Fund to Fight AIDS, Tuberculosis and Malaria⁵.

Confounded by politics, the debate has largely ignored the organizational exigencies of what Ernst-Ludwig Winnacker — president of the DFG, Germany's main research-funding agency — has called the “small science” of global health research⁶. ‘Big’ science requires initial down payments and large investments in infrastructure, but investment in ‘small’ science can be scattered among many funders engaged in an independent but coordinated pursuit of scientific inquiry. The happy deduction, as George Radda, head of the UK Medical Research Council, has noted⁷, is that there is no theoretical justification for preferring

a highly orchestrated, centrally controlled effort to a decentralized one.

Proposals

To initiate serious discussion, the CMH proposed the creation of a US\$1.5-billion fund to promote global health research, with a research model based on that seen in the organization of national health-research agencies and councils in North America, Europe and elsewhere. The rationale behind this proposal was that research agencies such as the US National Institutes of Health (NIH) and the UK Medical Research Council have long, proven track records in promoting the ‘best’ science in health R&D, through an investigator-initiated, peer-reviewed, merit-based research support system. Implicit in the CMH report was the concept that the global effort should tap into this enormous wealth of experience and focus on increasing the amount of research oriented towards improving the health of the poor in developing countries.

Although it is unlikely that the CMH's vision for a new global health-research fund will take root and flourish in the current economic and political climate, we can still explore the idea of harnessing the power of the international biomedical research community to respond to the most urgent of today's global health challenges. Although much of the existing public global R&D capacity is located in Australia, Japan, Europe and North America, a number of middle-

income countries — such as Brazil, Mexico, Argentina, South Africa, Malaysia, China, India and Thailand — are already investing in science, have developed similar mechanisms to promote excellence in research, and already contribute significantly to the global knowledge base. How could the skills, talents and resources of these scientific communities be put to better use in the service of improving health globally?

National agencies for biomedical and health research have at least three kinds of resources at their disposal that — if effectively harnessed — could make a difference in the near term. The first is the scientific capacity represented by their 'in-house' expertise and extensive networks linked to universities and independent research institutions.

Second, they determine their own scientific priorities and allocation of research resources, a portion of which is already used to fund initiatives in global health research. The amounts invested are substantial. Overall, governments in developed countries invested US\$34.2 billion in health R&D in 1998, with five countries (the United States, Japan, Germany, France and Britain) accounting for approximately 90% of the total⁶.

Third, the national research institutions for health and biomedicine have well-developed systems for funding and promoting the best science, relatively unfettered and unimpeded by politics, fostering rapid scientific discovery and advancement. In the United States, for example, these highly competitive funding mechanisms are governed by an independent and objective two-level peer-review system conducted by and for scientists. Expert peer review is part of the culture of science and is well integrated into NIH programmes that emphasize investigator-initiated proposals. Other programmes, including targeted planning awards and research programmes initiated by the agency, contracts for the research and method development that are considered necessary to stimulate scientific progress, and cooperative agreements with the private sector for product development are also subject to rigorous peer review.

After many discussions with directors and heads of the international programmes of the main biomedical agencies in North America and Europe, we believe that the time is now right to consider the creation of a global health-research network. Although many types of models are possible, the most feasible at the moment is a 'virtual' model. This consists of several of the world's leading medical research councils in a voluntary, independent but coordinated collaborative effort. Criteria for participation — beyond the commitment of their directors and senior staff — could include an assured national source of funding, scientific autonomy in priority setting and allocation of research resources, an independent peer-review system, a long and consistent

The international community must embrace the skills and expertise of any and all potential investors.

history of scientific accomplishment and engagement in science for global health. Generation of new resources would be nice but not necessary initially, as these research agencies could prioritize global health research when allocating resources that they already receive from their national governments, on the basis of scientific opportunity, disease-burden considerations, and potential impact. No single agency could — or should — be expected to shoulder this responsibility alone, yet together their investment could be significant. The major problem would be to keep the network coordinated, functional and focused on science rather than becoming a politically correct, all-inclusive, but dysfunctional, tool for development.

Not everyone will be satisfied with the narrow scope implied by this concept proposal. Some will correctly argue that an initiative by the biomedical research agencies in the scientifically developed high- and middle-income countries will not directly address a parallel problem of major concern to low-income developing countries — that of building the capacity of their own researchers and institutions. As the UN Secretary-General Kofi Annan puts it: "The idea of two worlds of science is anathema to the scientific spirit."⁹ Others may also argue, possibly correctly, that the involvement of the leading biomedical research agencies could inevitably encourage too much focus on basic science, and not enough in the areas of operations and health-services research. However, the network itself could address both issues, perhaps in conjunction with WHO and, eventually, a global research fund to support the development of science in the most resource-constrained nations.

Seizing the opportunity

To focus on such limitations is to ignore the urgency of the global health crisis, which has reached such enormous proportions that we cannot afford to wait for the perfect solution. Neither can we wait for the creation of a new fund for research while the existing Global Fund to Fight AIDS, Tuberculosis and Malaria (which is not oriented towards research) continues to search for adequate financial support, the most appropriate disbursement mechanisms and to demonstrate that it will actually make a difference in the absence of new tools (the products of new research) in controlling these devastating diseases.

Nor can we afford to permit large pockets of potential capacity to remain untapped. A vigorous response to the crisis is urgently needed now, and the international community must flexibly embrace the diverse range of skills, expertise and contributions of any and all potential investors.

When a coordinating council of the leading research agencies, north and south, is up and running, it will be a constant reminder to the heads of these agencies to remain committed and engaged in the global health-research effort, and will enable them to collectively identify the knowledge gaps that their agencies can address, alone or together. As they become better organized in coordinating, collaborating and complementing one another, it will become increasingly possible for development agencies and philanthropists to link their work — of capacity building in developing countries — to the basic and translational research growing out of the investments by the cooperating research agencies. Competition, an essential element to ensure excellence in science, can be focused on coming up with new knowledge in the shortest time frame, and not on which country or agency announces the discovery.

The alternative is to ignore the potentially large contributions of the national health-research councils and their effective mobilization. These nationally organized agencies have a specific mandate to focus on domestic health priorities, and so cannot easily play by the rules of other more globally oriented stakeholders — such as the multilateral agencies, foundations, or even the bilateral development agencies of their own countries. They are, after all, science agencies, and not donors. Yet while reflecting their own institutional cultures, they can make an important, highly significant contribution that would otherwise remain untapped. Given the enormity of the global health crisis we are facing, how long can we really afford to leave these institutions on the sidelines or only partially engaged? Not much, we submit — the time to move from concept to reality is now. ■

Gerald T. Keusch is director of the Fogarty International Center, National Institutes of Health, Bethesda Maryland 20892, USA. Carol A. Medlin is a Senior Researcher at the Institute for Global Health, University of California, San Francisco, California 94105, USA.

1. World Development Report: Investing in Health (World Bank, 1993).
2. Investing in Health Research and Development: Report of the Ad Hoc Committee on Health Research Relating to Future Intervention Options (World Health Organization, 1996).
3. Macroeconomics and Health: Investing in Health for Economic Development: Report of the Commission on Macroeconomics and Health (World Health Organization, 2001).
4. Feachem, R. & Medlin, C. *Nature* **417**, 695 (2002).
5. Nossal, G. J. *Immunol. Cell Biol.* **81**, 20–22 (2003).
6. Winnacker, E.-L. *Science* **295**, 446 (2002).
7. Radda, G. *Science* **295**, 445–446 (2002).
8. The 10/90 Report on Health Research, 2001–2002 (Global Forum for Health Research, 2002).
9. Annan, K. *Science* **299**, 1485 (2003).

Einstein not yet displaced

Controversial theory of varying speed of light still lacks a solid foundation.

**Faster Than the Speed of Light:
The Story of a Scientific
Speculation**

by João Magueijo

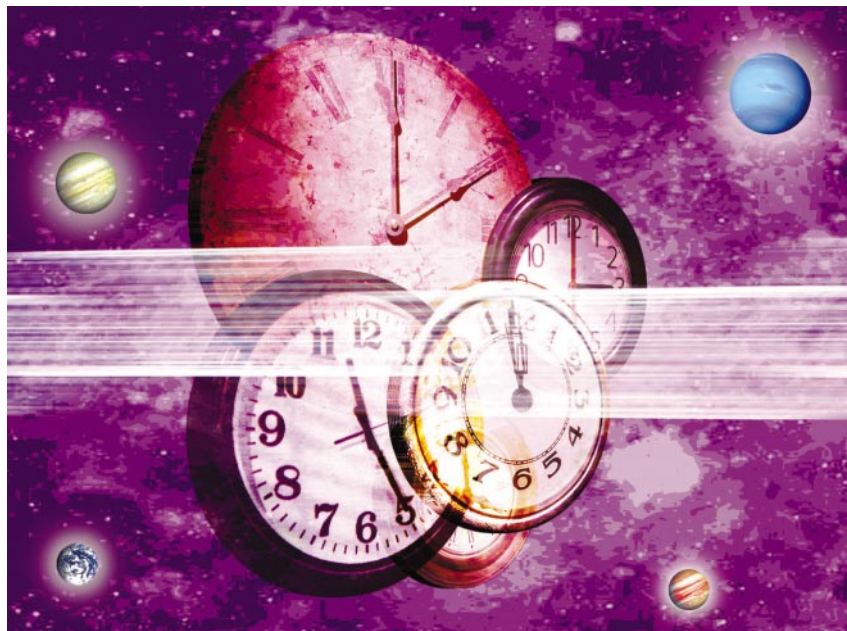
William Heinemann/Perseus: 2003. 320 pp.
£16.99/\$26

George Ellis

João Magueijo is one of many who hope to see the epitaph 'Einstein was wrong, I was right' on their gravestone. He is a cosmologist who, one rainy morning in Cambridge, suddenly saw the possibility of a varying speed of light (VSL) as an alternative to the inflationary-theory paradigm that dominates present-day theoretical cosmology. He knew from the start that it represented a fundamental challenge to physics orthodoxy (it violates the foundations of Einstein's special theory of relativity) and would not easily be accepted, but he worked enthusiastically to develop the idea. He found a collaborator who wavered but eventually completed a joint paper with him on the topic. This was rejected by major journals but was eventually accepted for publication after a long battle. He then discovered that the idea had already been proposed, in a slightly different form, by John Moffat. He found new collaborators and with them developed variants of his theory.

Faster Than the Speed of Light is a lively book that captures the excitement and frustrations of doing real-world science. Magueijo relates interestingly how his VSL proposal might possibly be a way out of some major puzzles facing cosmology, which he explains well. There are irritating passages, however, where he makes extended use of a metaphor involving farmers and cows in explaining relativity theory. Magueijo states that this is based on a dream that Einstein had as a boy — a fictional invention that displays such a cavalier attitude to historical truth as to call into question his other historical claims (and for the record, it was Richard Tolman, not Yakov Zeldovich, who first investigated the thermodynamics of bouncing universes). And at times Magueijo descends to an altogether different space characterized by hostile ranting ("seem to fancy themselves as scientific pimps") and crude language. In these passages he expresses his profound dissatisfaction with how he has been treated by the scientific world despite the recognition and generous support he has received (he was awarded a Cambridge fellowship and a Royal Society research fellowship, and is a reader at Imperial College, London).

His papers on VSL have now been widely



read and referred to. Why, then, his major discontent? He has had no more difficulty than many others who have presented challenges to orthodoxy. All major new ideas have been resisted in their time: the expanding Universe, continental drift, special relativity and quantum theory, for example. Science is inherently conservative — it has to be so, given the flood of speculative writing. It also has to be open, allowing dissemination of unorthodox views, which does occur. It is resistive but not impermeable, as is shown in his own case. There is a valid complaint, nevertheless: the current use of refereeing as a defence of the inflationary-theory orthodoxy in cosmology is indeed regrettable.

Magueijo's dissatisfaction is wider than that, however. He criticizes all university administration as parasitic and unnecessary, throwing in gratuitous insults as he does so. He is breathtakingly arrogant as regards funding — he seems to assume it is his right to be funded for the work he is doing with no questions asked. He gives no attention to the methods by which one can decide how public funding should be dispensed in science, nor to why the public should pay any money at all to people like him. Yes, there are problems in university organization and the funding system; constructive criticism is justifiable and indeed needed. But his remarks are purely destructive.

What of the VSL theory itself? Is it the panacea he hopes for? No, it is not. Einstein reflected deeply on the foundations of physics, and that was the basis of his success.

Magueijo has not gone back to the foundations and sorted them out. Any theory of this kind needs, first, a viable proposal for measuring both time and distance, as velocity is based on this; second, a physical model that embodies the results of these measurements in some well-defined mathematical structure; and third, a theory of electromagnetism that predicts the speed of light in relation to these measurement processes. He has none of these, and without them he does not have the basis to put his theory on a solid foundation.

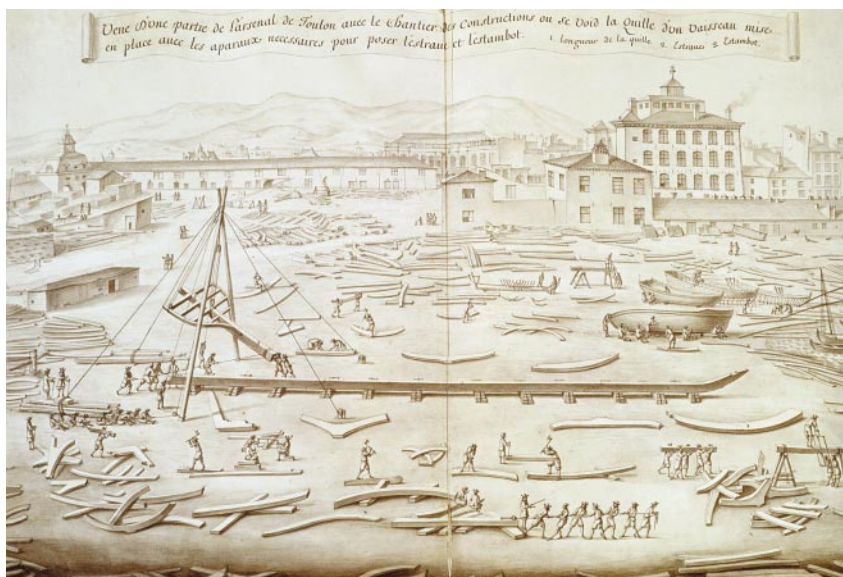
Standard relativity theory deals with all these issues in depth. The key point is that current ways of measuring distance precisely incorporate the speed of light in their foundations. On large scales, radar (with its variants such as the Global Positioning System) is the only viable method. It is then not possible for the speed of light to vary, because it is the very basis of measuring distance; as emphasized by J. L. Synge, the natural units for distance are light seconds or light years, rather than metres or miles. Furthermore, this is then built into the foundations of the theory through the space-time metric tensor and its interpretation as determining proper time (time measured by an ideal clock along its worldline), proper distance (measured by radar), and the null cone (characterizing the path of light through space-time). Because Magueijo and Moffat ignore this physical interpretation of the metric, their so-called 'phase transition in the speed of light' is just a jump in arbitrary

units for time, unrelated to measurement procedures. It is not a physical prediction.

Furthermore, the variation principles proposed as underlying the physics involve the metric tensor in raising and lowering indices to create scalars — and hence build into the foundations of the theory the invariance of the speed of light (the metric determines the speed of wave propagation). We are given no reason why any broken symmetries associated with special solutions of the resulting equations will give a causal explanation for a varying speed of light — but this variation is the arbitrary postulate of VSL theory. And apart from the part of the action determining variation of the speed of light (independently of Maxwell's equations), the explicit occurrence of the speed of light in the VSL variational principle proposed is only in a ratio with the gravitational constant G — so this is just a varying- G theory in disguise.

Developments that could make VSL viable, such as further investigation of the time variation of the fine-structure constant, of two-metric theories, of an altered version of the symmetry group underlying relativity theory, or through a string-theory motivation for varying 'constants', need to provide a clear relation to space and time measurement, as well as a physical reason (based in some version of Maxwell's equations) for the speed of light to vary. It is a pity that Magueijo does not mention progress made in these directions by workers other than himself and his own collaborators. ■

George Ellis is in the Department of Mathematics, University of Cape Town, Rondebosch 7700, Cape Town, South Africa.



More than the sum of its parts: the relationship between the planks determines what kind of boat it is.

C'est la vie?

The Delphic Boat: What Genomes Tell Us

by Antoine Danchin, transl. Alison Quayle
Harvard University Press: 2003. 368 pp. \$35, £23.50, E35

Axel Meyer

This contribution from old Europe's *grande nation* discusses genomics, the politics of genome research, the philosophy of science, and the not-so-small question of the nature

of life. The first problem faced by the author, Antoine Danchin of the Pasteur Institute in Paris, France, is how best to describe a genome. He borrows from Greek mythology the tale of the oracle of Delphi, which asked whether a boat that has had all of its planks replaced over time is still the original boat or not (to its owner, who watched its evolution, it would be). Danchin makes the point that it is the relationship of the planks (or in the case of the genome, the genes) to each other that determines what kind of boat (genome) it is; individual planks are less important in determining the essence of a boat.

It has been known for some time that a genome is not merely a set of independent genes arranged like pearls on a string. The essence of a genome has been described as a code, a blueprint, a musical score and a set of instructions. All of these metaphors are used in an attempt to convey the notion that a genome is more than the sum of its parts.

From an evolutionary biologist's point of view, each organism's genome (including our own) is a record of its evolutionary history: genomes are shaped by symbioses and hybridizations, as well as by natural selection. In attempting to understand the origin and diversification of life, the increase in complexity during ontogenetic development, and even more straightforward questions such as the genetic bases of diseases, the study of individual genes will reach its limits and fall short of a more holistic appreciation that considers the entire genome (and the entire phenotype). Genes 'talk' to one another, regulating each other's expression in response to environmental conditions and the prevailing ontogenetic or metabolic state of the cell or organism. Genes are also affected by their position in the chromosome, by the base composition

Preying for dinner-time

Mantids are ambush predators that keep still and wait for other insects to approach before striking and feeding on them. Like other mantids, the female mantis *Harpagomantis discolor* shown here is camouflaged and remains motionless, or rocks slightly from side to side as if swaying in the breeze, all the while keeping its forelegs folded in a manner reminiscent of prayer. When another insect comes close, the mantis snatches it and holds it in a pincer-like grip while devouring it alive. Mantids are just one of the arthropod orders included in *The New Encyclopedia of Insects and their Allies* (Oxford University Press, £25), edited by Christopher O'Toole.



of surrounding DNA, and by other “local climates”, as Danchin puts it, that might influence their tempo and mode of evolution or level of expression.

Early and overly simplistic notions that the determination of a genome sequence (or “genome text”, in Danchin’s words) will determine precisely how an organism works have long gone the way of the dodo. *In silico* analyses — computer analyses of genomic information as an alternative to *in vivo* or *in vitro* studies — are now an established field, known as bioinformatics. They are largely based on comparative (evolutionary) approaches. The hardest route, but in my opinion the one that offers the brightest future and the most intellectual profit, might be comparative and functional-genomic analyses of the architecture of genetic cascades and networks. The hope, and there are encouraging signs for this, is that there are commonalities and higher-level organizational principles, or even that this field of enquiry might reveal some of biology’s more elusive laws.

Of course, no book can be written for everybody, but many readers might like to know what, if anything, genomics can tell us about the meaning of life. Various features of this book will be of varying interest to different audiences. Given the big issues it addresses, *The Delphic Boat* is apparently intended for a wide, non-specialist readership. Descriptions of the idiosyncrasies, foibles, imperfections and particular aspects of the politics of French grant-funding agencies make the book a lively read and demonstrate to the uninitiated reader that politics has an unpleasantly large role in science. Danchin also stresses the important contributions of French researchers to the various advances in genomics.

Although these often rather personal and historical sidelines might be of particular interest to science historians or researchers working on microbial genomics, these specialist readers would not want to be bothered with explanations of how the polymerase chain reaction works, for example. But such techniques and other introductory and background information will presumably be of interest to the non-specialist reader who would like to discover what genomics is all about. This kind of reader might need more explanation of some of the basic concepts in genetics and genomics, however. In some places, rather specific jargon is used which might confuse general readers. And discussion of alternative DNA-sequencing techniques is probably beyond most uninitiated readers. ‘Automatic’ sequencing techniques have been used pretty much exclusively in genomics for 15 years now, but one of the few figures in the book shows a ‘manual’ gel with radioactively labelled DNA — rather a relic of the past. The lay reader would have profited from the inclusion of



String theory? The Kronos Quartet perform *Sun Rings* against a backdrop of images of the Universe.

an electropherogram instead, if the budget had stretched to colour figures.

These limitations aside, Danchin provides an authoritative and impressively multidisciplinary treatment of many aspects of genomics, and his provocative thinking about the *raison d'être* captured my interest. I suspect that it will do the same for many readers from a large number of scientific disciplines, not only biological ones. ■

Axel Meyer is in the Department of Biology, University of Konstanz, 78457 Konstanz, Germany.

Music

Sounds from space

Sun Rings

by Terry Riley

Performed by the Kronos Quartet

Juliane C. Mössinger

During the First World War, the German physicist Heinrich Barkhausen accidentally made the first recording of strange ‘whistling’ sounds while tapping British phones. He reported the peculiar tones (*Phys. Z.* **20**, 401; 1919), but could not explain them at the time.

The ‘whistler’ is just one of many space sounds that form the inspiration for Terry Riley’s composition *Sun Rings*, performed by the Kronos Quartet at the Barbican Centre in London on 22 March 2003, with further performances to follow in seven US cities over the coming months. The work, which was commissioned by NASA’s arts programme, is based on the collected space recordings of Don Gurnett, an astrophysicist at Iowa State University, whose plasma-wave receivers have been travelling

on various spacecraft for the past 40 years.

Plasma waves are in low radio frequencies. When converted to audible sound waves, patterns emerge — like the ‘whistler’, a rapidly descending tone caused by lightning discharges. Another is the ‘dawn chorus’, brief, rising-frequency tones that sound like the chirping of birds and which are produced by electrons trapped in magnetic fields surrounding planets.

For the performance, the theatre is plunged into darkness. The quartet on stage is surrounded by thin silver sticks adorned with dimly glowing, star-like lamps. Behind the musicians, digital images of edited and original NASA footage, arranged by visual designer Willie Williams, are projected onto a giant screen.

The ten ‘spacescapes’ that make up *Sun Rings* incorporate the original sounds recorded by the plasma-wave receivers. The string quartet also mimics them. But Riley’s interesting and diverse composition is much more than this. He has done what scientists consider unthinkable, explained Gurnett in his pre-performance talk: he has manipulated original data and changed them beyond recognition. Riley’s journey is largely about humans as they explore outer space to gain an awareness of their Solar System neighbourhood. The human dimension is emphasized by the addition of a choir in some of the movements.

‘One Earth, one people, one love,’ a female voice declares. Images about life on Earth are displayed on the huge screen behind the Kronos Quartet. They are part of the message sent on Voyager through space. While bombs are dropping on Iraq, ‘one Earth, one people, one love’ seems to be further away from humanity than Voyager could ever travel. ■

Juliane C. Mössinger is a physics editor at Nature.

Pipe dreams

What was your first experiment as a child?

My chief technician at the time, my younger sister Gail, remembers our studies on crayfish: "We would catch the crusty little bug-gers by dangling pieces of hot-dog on the end of a string. When they pinched, we'd toss them in a bucket and bring them home, and Mom would make us leave them overnight on the porch. Next day we'd dump them back in the pond and start all over again. We must have caught the same dozen crayfish hundreds of times. I imagine a whole crayfish mythology has arisen around those mysterious disappearances and reappearances. Like people who are abducted by aliens."

Who has been the most important mentor in your career?

This is an incredibly difficult question. I have been extremely fortunate in having had a series of excellent mentors at nearly all stages in my career. During the past few years, I have given seminars at my high school and colleges, and many of my mentors have attended. As I mature, I realize how important it is to thank them as they have given so much of themselves to help me achieve.

Whose graduate student would you most like to have been (historical impossibility notwithstanding)?

Maybe T. H. Huxley. But the supervisor/student relationship is like a medieval apprenticeship, and the interpersonal chemistry is hard to predict. And would he have taken a female student?

What book has been most influential in your scientific career?

Books? Aren't those the things that provide insulation in one's office and decorative accents in one's home?

What literary character would you employ as a postdoc?

Daedalus. He had family problems, of course, but hopefully they wouldn't interfere.

Which field in science (apart from your own) deserves more funding, and why?

High-risk/high-pay-off research. Funding agencies are far too conservative, so the most creative ideas may never get beyond the stage of dreams.

You have the audience in your hands, but some smart-alec asks you the killer question you have no idea how to answer. What's your response?

The wonderful thing about being over 40 is being able to say: "I don't know. I'm so glad you brought that up. Let me think about it."

What book currently resides on your bedside table?

J. R. R. Tolkien's *The Hobbit*, *The Crucible of Creation* by Simon Conway Morris and *Sparks of Life* by James Strick. I am dreadfully behind.

What music heads the playlist in your car or laboratory?

The wonderful thing about the iPod is that I can give you an answer to the megabyte. So, in megabyte order: Celtic music including a hefty amount of bagpipe; sea shanties and other folk music; baroque and early music; classical; Gilbert and Sullivan, musicals, ragtime and jazz.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Benjamin Franklin. He had it all: charm, scientific insights, and was at the nexus of world politics. And, by all accounts, he was quite a ladies' man.

What one thing would you rescue from your burning laboratory?

Right now I am blessed with two outstanding technicians. Naturally I have told each that they would be it.

What would you have become, if not a scientist?

Probably a historian of science, or an environmental lawyer. I briefly flirted with each, and was told by the relevant professors that I would miss doing science too much. They were, of course, right.

What previously unrecognized sport or pastime should be included in the Olympic Games?

The contestants are given a bag of spinach and asked to isolate the DNA, purify the ribulose-1,5-bisphosphate carboxylase, and separate the rRNA from the mRNA. This would be a timed trial with certain as-yet-undetermined constraints, but leaning towards the low-tech.

Do you have a burning ambition to do or learn something of no practical or immediate value?

I have been learning bagpipes. But it is of practical and immediate value: it clears my mind and house, and gives me pleasure.

You've just been told (in confidence) that the world will end tomorrow. What would you do next?

See if there was anything I could do to get off with my family, as in *The Hitchhikers' Guide to the Galaxy*.

Harry Potter or Lord of the Rings?

Harry Potter. If I don't give this answer my



Lynn Rothschild

Lynn Rothschild is an evolutionary biologist at the NASA Ames Research Center and is currently immersed in the field of astrobiology. She is president of the Society of Protozoologists and is interested in how microbes survive and evolve in the physical environment on Earth and potentially elsewhere.

ten-year-old will break off all diplomatic ties with me, even if I am his mother. However, I have now seen *Lord of the Rings* and it should be rated PG50. But, in fairness, my parents did warn me.

Due to an unexpected change in Earth's rotation, days now last 25 hours. How will you spend the extra hour?

Ah, but it will happen due to the frictional interaction between Earth and our dear Moon. But sadly not in our lifetime. No doubt I would spend it bemoaning the fact that we weren't up to 26 hours yet.

The host at a dinner party has placed you next to God. What would you talk about?

Hmmm. A type 'A' personality with an obvious sense of humour. This would be my plan on going in:

Appetizer: Why are you talking to me?

Main course: What is your role? How did life originate and evolve? Are we alone in the Universe(s)?

Sorbet: What happens after death?

Dessert: What is your take on religion?

After-dinner drinks: What do you want me to do now?

You're the only scientist at a party. How do you describe what you do for a living?

"Investment banking? Really? How nice. I look for life on Mars."

What's your motto?

Omniscience tempered by amnesia.

Pebbles in the nodal pond

Jan Zaanen

Rippling patterns of electron waves in a copper oxide match the expectation for a certain kind of excitation — another step towards understanding why copper oxides superconduct at far higher temperatures than other materials.

Perhaps the first image that comes to mind when thinking about waves is the interference pattern on the surface of a pond after some pebbles are thrown in. According to quantum physics, electrons can behave like waves, and so give rise to similar phenomena: the 'pebbles' are imperfections in the medium, scattering the electron waves, and the 'ripples', or interference fringes, can be seen through a scanning tunnelling microscope (STM). On page 592 of this issue, McElroy *et al.*¹ show that such patterns can be used to study the details of a special kind of electron wave — so-called nodal fermions — that occurs in high-temperature superconductors.

Scanning tunnelling microscopes are remarkable machines. Tuned the right way, they can image the wavefunctions (essentially, the probability distributions) of electrons directly. Beautiful images have been captured of electron waves, for example, in the vicinity of imperfections on the surfaces of simple metals (Fig. 1a), and surrounding artificially tailored nanostructures such as 'quantum corrals' (see, for example, ref. 2). These phenomena have long been understood. Far more exciting from the perspective of fundamental physics is the application of STM imaging in the context of high-temperature superconductivity.

Although discovered more than 15 years ago, high-temperature superconductors continue to fascinate physicists. They are copper oxides that superconduct — that is, offer no resistance to the flow of electric current — at unusually high temperatures (typically 100 K), and their electrons behave very differently from those in simple metals³. It seems that there are unusually strong interactions between the electrons in a high-temperature superconductor, but the full story is still a mystery.

Images from an STM, like those recorded by McElroy *et al.*¹, would show wave-like patterns if the high-temperature superconductors behaved as simple metals. But in fact the real-space maps look completely different. Instead of a 'pebbles-in-a-pond' pattern, they show a pattern composed of irregular but basically straight lines, forming a more or less orderly texture^{4,5}, like fabric woven from raw silk (Fig. 1b). When images such as these were first seen⁵, many experts took them to be evidence for 'stripes'. Stripes are a form of order in copper oxides, known

to compete with superconductivity, that cause the electrons to come to a standstill in regular patterns. Although stripe signals might be hidden in the data^{5,6}, it is now clear that most of what is seen has nothing to do with stripes. Instead, the raw-silk texture is caused by electron waves being scattered by imperfections. But these waves are very different from the 'pebbles-in-a-pond' waves in simple metals — they are 'nodal fermions', special electron waves associated with the unconventional nature of the superconducting state in copper oxides.

If the raw-silk patterns are associated with nodal waves, they should be composed of a total of 16 different modulations, each having a unique dependence on the energy of the scattered electrons. Using Fourier analysis to nail down the precise properties of the waves, McElroy *et al.* have identified all of these modulations in their STM data and have reconstructed in great detail the behaviour of the underlying nodal fermions. Their results confirm earlier observations of these nodal states in photoemission experiments, but go far beyond them in terms of the detail revealed.

The existence of nodal fermions is implied by the 45-year-old Bardeen–Cooper–Schrieffer theory of conventional superconductivity (extended to include *d*-wave symmetry). So their detection by McElroy *et al.* might seem unexciting. But in fact there is more to it. This work is an experimental innovation that has opened up a new window of observation on the mystery of high-temperature superconductivity. Although the nodal waves might seem at first sight to be quite conventional, there is evidence suggesting that they are much more fragile than the electron waves in normal metals. It appears that they behave like quantum waves only when the system as a whole is in a macroscopic quantum state — that is, when it is superconducting⁷.

McElroy and colleagues' experiments have sprung a great surprise, showing just how different these excitations are: when the electron energy approaches that of the superconducting gap (associated with the short timescale on which the physics giving rise to superconductivity originates), the STM patterns change in a sudden and dramatic way. The raw-silk texture disappears, and is replaced by a very different pattern (Fig. 1c)^{1,8}. I call these patterns 'quantum salad-dressing'⁹: there seem to be two very different states of electron matter present

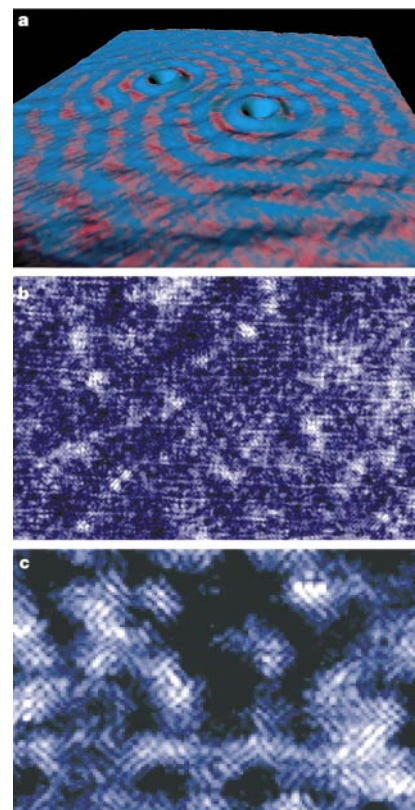


Figure 1 Electron waves. With a scanning tunnelling microscope, quantum-mechanical wave patterns can be observed directly: when electrons in conventional metals such as copper scatter off imperfections in the crystal lattice (a), a pattern of ripples forms². McElroy *et al.*¹ have similarly detected interference patterns in a high-temperature copper-oxide superconductor (b). At first sight, these do not look like waves, but more like fabric woven from raw silk. In fact, these patterns are caused by special electron waves, called nodal fermions, that occur in high-temperature superconductors. McElroy *et al.* also show that, contrary to theoretical expectations, the patterns suddenly change when the electron energy is raised (c), shedding new light on the mysterious nature of electron states in copper-oxide superconductors.

that, like oil and vinegar, do not want to mix. These findings suggest that, at high energies, the electron system reveals its true nature as a strongly interacting quantum fluid, which is expected to bear some similarities to a classical fluid. Only when the energy becomes

low enough does the quantum coherence of the superconducting ground state take charge, and excitations then act like 'simple' quantum-mechanical waves.

The great potential of this technique lies in its unique sensitivity to the coherent, wave-like nature of the excitations. It should now be possible to investigate some of the burning issues in the field. For instance, the interplay between stripes and nodal fermions is a contentious issue. Although the presence of a very weak remnant of stripe order in the best superconductors is still debated^{1,6}, stripes usually appear only when superconductivity is suppressed. Magnetic fields are the natural enemy of superconductivity: patches of stripe-like order around lines of concentrated magnetic flux (vortices) have already been detected in STM images^{10,11}. Do stripes destroy the nodal fermions? Nobody knows, but using STM imaging it should be possible to take a direct look.

An intriguing question is what happens to the raw-silk texture when the temperature rises. There are indications that increasing temperature is detrimental to nodal

excitations, but so far the evidence is rather indirect⁷. The capacity of the STM to probe quantum coherence is unique, and I anticipate spectacular results in this area in the near future. If our present understanding is correct, the raw-silk texture should disappear quickly with rising temperature, telling an exciting story about the delicate role of quantum physics in this strange electron system. ■

Jan Zaanen is at the Instituut Lorentz for Theoretical Physics, Leiden University, PO Box 9506, 2300 RA Leiden, The Netherlands. e-mail: jan@lorentz.leidenuniv.nl

1. McElroy, K. *et al.* *Nature* **422**, 592–596 (2003).
2. www.almaden.ibm.com/vis/stm/stm.html
3. Orenstein, J. & Millis, A. J. *Science* **288**, 468–474 (2000).
4. Hoffman, J. E. *et al.* *Science* **297**, 1148–1151 (2002).
5. Howald, C., Eisaki, H., Kaneko, N. & Kapitulnik, A. *Proc. Natl Acad. Sci.* (in the press); preprint cond-mat/0201546 at <<http://arXiv.org>> (2002).
6. Howald, C., Fournier, P. & Kapitulnik, A. *Phys. Rev. B* **67**, 014533 (2002).
7. Feng, D. L. *et al.* *Science* **289**, 277–281 (2000).
8. Lang, K. M. *et al.* *Nature* **415**, 412–416 (2002).
9. Zaanen, J. *Nature* **415**, 379–381 (2002).
10. Hofmann, J. E. *et al.* *Science* **295**, 466–469 (2002).
11. Sachdev, S. & Zhang, S. C. *Science* **295**, 452–454 (2002).

Developmental biology

How neurons avoid derailment

Paul A. Garrity

During development, neurons extend thin protrusions that must choose between alternative routes. A study of this process in fruitflies unites two previously disparate protein families.

In many ways, the developing nervous system resembles a city at rush hour, with large numbers of neurons — up to 10^{12} in humans — extending thin protrusions called axons that take highly specific routes to reach their destinations. This ability of axons to choose one particular path from many alternatives is essential for precise wiring of the nervous system. On page 583 of this issue, Yoshikawa and colleagues¹ present the next chapter in the unfolding story of how axons find their way.

The central nervous system of the fruitfly (*Drosophila melanogaster*) embryo provides a simple example of pathway selection, as axons growing across the midline of the nervous system — thereby connecting the left and right halves of the animal — choose between two alternative routes. Roughly half the axons in each body segment choose the anterior route, establishing an anterior axon bundle (commissure); the other half take the posterior path, establishing a posterior commissure.

Ideas about the mechanism behind this decision began to take shape in 1999, from work by Bonkowski *et al.*² on Derailed, a member of the well-studied RYK family of receptor proteins that span the cell

membrane. They found that Derailed is both necessary and sufficient to direct axons through the anterior commissure. It is expressed specifically on axons that choose the anterior route; without this protein, the axons take a variety of paths across the midline, often wandering between the commissures. Moreover, if Derailed is experimentally misexpressed in neurons that normally send axons through the posterior route, these axons switch to the anterior path.

To further explore how Derailed works, Bonkowski *et al.* used a soluble, labelled version of the extracellular portion of the protein as a probe on tissue. This detects any cell-surface binding partners (ligands) for Derailed. (In fact, it appears to measure the distribution of any 'free' ligand — ligand that is not already bound to natural Derailed — a detail that will become more important below.) In this way, the authors detected a potential binding partner specifically in the posterior commissure. The implication was that Derailed directs axons into the anterior commissure by repelling them from the region of ligand expression in the posterior commissure. But the identity of the Derailed ligand (and the ligands of RYK-family receptors in general) remained unknown.

Yoshikawa *et al.*¹ now fill this gap in our knowledge of the Derailed ligand, with the revelation that it belongs to the Wnt family of secreted signalling proteins. Wnt proteins have long been known to be essential in animal development: they regulate cell-fate determination, cell proliferation and movement, and tissue polarity^{3,4}. Inappropriate activation of Wnt signalling leads to colorectal and other cancers in humans⁵. Moreover, Wnts regulate the formation of connections between neurons and their targets (other neurons or muscle)^{6,7}. The signalling pathways through which Wnts control cell fate and tissue polarity have been studied in depth, but how they exert their effects on neuronal projections has largely been a mystery. By showing that *Drosophila* Wnt5 functions through Derailed, Yoshikawa *et al.* uncover a role for Wnt signalling in axon guidance. Their findings also reveal a connection between two previously disparate but large fields of research — the study of Wnt proteins and of RYK-family members.

Yoshikawa *et al.* were led to Wnt5 by using a clever genetic screen for regulators of Derailed activity. As Derailed misexpression switches axons that would normally project through the posterior commissure into the anterior commissure, the authors reasoned that reducing the function of genes that promote Derailed signalling might cause these axons to switch back. After systematically examining animals that had reduced function in many different genes, the authors found that reducing the activity of the *wnt5* gene lessened the ability of misexpressed Derailed to switch axons into the anterior commissure. Further genetic analysis showed that *wnt5* mutants have similar axon-guidance defects to *derailed* mutants, consistent with the idea that these genes work together during normal development. Moreover, overexpressing Wnt5 throughout the midline of otherwise normal animals completely prevented the anterior commissure from forming, whereas Wnt5 overexpression in *derailed* mutants had no such effect. These results imply that Wnt5 repels Derailed-expressing axons.

So, Wnt5 and Derailed function together to guide axons into the anterior commissure. But do these proteins actually interact? This seemed likely, given that the extracellular portion of Derailed contains a region that is related to a previously characterized Wnt-binding domain. To find out for sure, Yoshikawa *et al.* used the Derailed extracellular domain to probe *Drosophila* extracts, and found that this domain did indeed bind to Wnt5. Moreover, the soluble, labelled extracellular domain of Derailed (used previously² to probe for Derailed ligands) failed to bind to *wnt5* mutant embryos, but bound both commissures when embryos overexpressed Wnt5. So, genetic and biochemical data indicate that

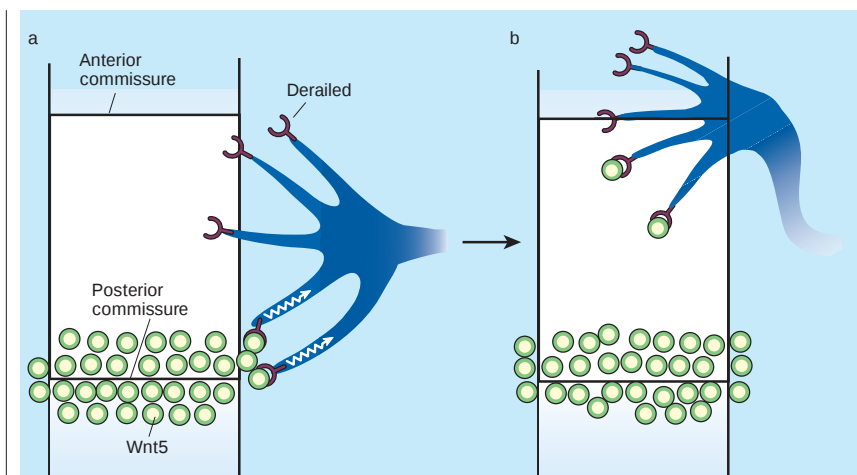


Figure 1 Directing axons in the developing fruitfly central nervous system. **a**, Axons that express the Derailed receptor protein are destined for the anterior commissure (axon bundle). Yoshikawa *et al.*¹ have found that Derailed binds to Wnt5 protein — which is present in a region surrounding the posterior commissure — sending signals (wavy arrows) into the neuron. **b**, This interaction appears to drive axons away from the posterior commissure and into the anterior one. The axons seem to retain the Wnt5–Derailed complex on their surface as they grow through the anterior commissure.

Wnt5 binds Derailed and prevents Derailed-expressing axons from entering the posterior commissure (Fig. 1).

The data also suggest that these proteins are subjected to additional levels of regulation. For instance, one might predict that Wnt5 protein would be present in the posterior commissure — from which it would repel Derailed-expressing axons — but absent from the anterior commissure. Yet Yoshikawa *et al.* show that, although Wnt5-encoding RNA is concentrated in cell bodies near the posterior commissure, the protein itself is present on both commissures. This contrasts with the previous finding² that the soluble Derailed extracellular domain binds specifically to the posterior commissure. Interestingly, Yoshikawa *et al.* found that soluble Derailed binds to both commissures of *derailed* mutants, again indicating that Wnt5 is present on both commissures. As mentioned above, using soluble Derailed appears to measure the distribution of its ligand that is not bound to natural Derailed. So it could be that, in normal fruitfly embryos, the Wnt5 in the anterior commissure is already bound to natural Derailed, preventing it from being detected by using soluble Derailed. In *derailed* mutants, all the Wnt5 might be 'free'.

Why, then, are Derailed-expressing axons that are exposed to Wnt5 in the anterior commissure not discouraged from midline crossing? Perhaps the Wnt5–Derailed complex signals only briefly, or the axons have a limited time window in which to decide whether to move away from or towards the Wnt source before other signals take over. Alternatively, co-factors might potentiate the repellent effect of Wnt5 in the posterior commissure, or silence the repellent effect in the anterior commissure.

Studies of Wnt proteins in cell-fate

determination and tissue polarity have uncovered a rich network comprising Wnt-detecting complexes on the surface of cells and intracellular signalling pathways that are regulated by these receptors^{3,4}. Some intracellular signalling molecules, such as Dishevelled proteins, seem to function downstream of Wnts in many situations, but the signalling network also appears to contain branch points, with some branches

directly affecting the cell's 'skeleton' and others affecting gene expression. It remains to be seen how these well-studied pathways contribute to Wnt5's role in axon guidance.

Conversely, it will be interesting to see whether RYK-family members contribute to Wnt signalling during cell-fate and tissue-polarity decisions and, if they do, how they interact with the Wnt-regulated pathways that control these choices. In particular, do RYK proteins act as subunits of more conventional Wnt-receptor complexes? Do their intracellular domains associate with proteins such as Dishevelled? Or do the RYK proteins function in other ways, for instance via enzymes such as Eph receptors, with which they interact in vertebrate cells⁵? The findings of Yoshikawa *et al.*¹ bring an important new player to the Wnt signalling system, and may represent only the first of many connections between these two protein families.

Paul A. Garrity is in the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

e-mail: pgarrity@mit.edu

1. Yoshikawa, S., McKinnon, R. D., Kokel, M. & Thomas, J. B. *Nature* **422**, 583–588 (2003).
2. Bonkowski, J. L., Yoshikawa, S., O'Keefe, D. D., Scully, A. L. & Thomas, J. B. *Nature* **402**, 540–544 (1999).
3. Wodarz, A. & Nusse, R. *Annu. Rev. Cell. Dev. Biol.* **14**, 59–88 (1998).
4. Nusse, R. www.stanford.edu/~rnusse/wntwindow.html
5. Bienz, M. & Clevers, H. *Cell* **103**, 311–320 (2000).
6. Hall, A. C., Lucas, F. R. & Salinas, P. C. *Cell* **100**, 525–535 (2000).
7. Packard, M. *et al.* *Cell* **111**, 319–330 (2002).
8. Halford, M. M. *et al.* *Nature Genet.* **25**, 414–418 (2000).

Mathematics

Regime change in meteorology

Ian Stewart

Sudden transitions between large-scale atmospheric circulation patterns — kinds of 'punctuated equilibria' — have a deterministic component that can be exploited to identify preferred dynamic cycles.

Weather is notoriously unpredictable. But there are many kinds of prediction, and it is possible to make scientific predictions about weather without predicting the weather. A case in point is the common occurrence of 'blocked' patterns of weather, in which the atmosphere remains in much the same state for a week or more, before suddenly switching to another long-lived pattern. In a thesis just completed, and in an earlier paper, Daan Cromelin^{1,2} shows how established work in theoretical mathematics can be brought to bear on these phenomena.

Examples of atmospheric pattern-switching include the North Atlantic Oscillation (NAO), the Arctic Oscillation, and the Pacific–North American pattern. Each of these is a specific instance of variability in the large-scale atmospheric circulation. For

example, during the positive phase of the NAO, the winds over the North Atlantic are anomalously westerly; during the negative phase they are anomalously easterly. When the atmospheric circulation is dominated by one or other of these phases, the weather itself (precipitation and temperature) can be dramatic. For example, winter flooding in northern Europe often accompanies the positive phase of the NAO, whereas the negative phase is usually characterized by severe cold. Much is known about these circulation patterns, but understanding the processes associated with transitions between them — which are perhaps their most significant feature — is another matter entirely.

In 1999, Palmer³ showed that the non-linear dynamics associated with these circulation patterns affects the atmosphere's response to some imposed change — for

instance, increased levels of carbon dioxide, the gas mainly responsible for the anthropogenically induced greenhouse effect. To understand global warming and its implications for climate in general, then, we first need to understand the dynamics of atmospheric regimes and regime change.

In theoretical work supported by observational data^{1,2}, Cromellin has followed Palmer's lead and provided evidence that transitions between atmospheric regimes may be related to the occurrence of a 'heteroclinic cycle' in nonlinear models of large-scale atmospheric dynamics. Heteroclinic cycles explain both the apparent stability of the blocked regimes and the sudden changes that occur when they are unblocked. The usual statistical approach views transitions between regimes as random processes, but there has been growing evidence that atmospheric regime change has a deterministic component. Cromellin's work pins this proposal down and makes it possible to test it.

Heteroclinic cycles have an element of unpredictability, but their dynamics is relatively straightforward and much of it is very predictable indeed. The cycles are characterized by occasional bursts of activity punctuated by lengthy periods of torpor. Specifically, heteroclinic cycles are sequences of connections between unstable equilibrium states — 'saddlepoints' that are stable in some directions but unstable in others. The torpid states are predictable: they occur when the system is near an equilibrium. The uncertainty lies in predicting when the torpor will cease.

For example, a pendulum has an unstable equilibrium in which it balances vertically in an upward position. If disturbed, the pendulum rotates with increasing speed to the

downward vertical, and then climbs back to the upward vertical. In the absence of friction, it returns very close to the original unstable equilibrium, but approaches it from the opposite side. This is a connection between an equilibrium and itself — in the jargon, a 'homoclinic' connection.

Connections are also possible between distinct equilibria, but these are uncommon unless the system possesses a degree of symmetry. In the case of the Earth and its atmosphere, there are two kinds of symmetry. Thus any model of the system is mathematically unchanged by rotation about the same axis, or if north and south are flipped by reflection in a mirror placed in the equatorial plane. These symmetries, broken only slightly by the effect of uneven landmass distribution, can create complicated patterns of heteroclinic connections. Mathematicians have carried out much theoretical work on heteroclinic cycles; now we are starting to see the fruits of this research in applied science.

Cromellin begins by analysing 'empirical eigenfunctions', or common flow patterns, in the atmosphere. In this technique, the actual flow is approximated by the closest possible combination of a set of independent basic patterns. The complicated equations of meteorology are thereby transformed into a system of ordinary differential equations describing a finite number of variables, which represent the amplitudes of the component patterns. Heteroclinic connections can be interpreted as the persistence of various apparently stable flow patterns, punctuated by rapid switches from one such pattern to another. This apparently counterintuitive behaviour makes perfect sense in the context of a heteroclinic cycle.

Cromellin tests his theory using data on the Northern Hemisphere for the period 1948–2000. He computes empirical eigenfunctions, and uses them to approximate the weather patterns. Next, he determines the most probable patterns, which should correspond to possible equilibria in any underlying deterministic model; six of these patterns dominate the statistics. Transitions between these six patterns are not purely random, but display some systematic features, which can be described using a Markov chain — a matrix of transition probabilities. Finally, the underlying deterministic dynamics can be inferred.

This analysis demonstrates the existence of a preferred dynamic cycle connecting various regimes of blocked flow in the Atlantic region. The dynamical variable is the deviation of the atmospheric flow from its mean state, which we will refer to simply as the 'flow'. The cycle begins with north-south flows over the Pacific and North Atlantic. These merge to form a single east-west Arctic flow. This in turn elongates over Eurasia and the west coast of North America, leading to a predominantly north-south flow. In the second half of the cycle, the flow patterns revert to their original state, but following a different sequence. The complete cycle takes about 20 days. The details of the cycle suggest that, in effect, the North Atlantic and Arctic Oscillations may act as a partial trigger for each other.

Cromellin's first contribution is to provide a systematic technique for making the preferred regimes visible from real data. His second is to show that, on an appropriate scale, transitions between atmospheric circulation regimes may have an underlying

Earth science

Stressed in Alaska

The seismic events in central Alaska on 23 October last year, and then on 3 November when a magnitude 7.9 earthquake ruptured more than 300 km of the Earth's surface, won't rate much mention in human history. Such was the remoteness of the area that, happily, no one was killed. And despite the forces involved, as seen in this image of ground movement along a highway, and in the occurrence of many huge landslides, damage to infrastructure was comparatively light.

Now, however, analyses of the earthquakes are being recorded for scientific history. One of the first reports appears in *Geophysical Research Letters*



(doi:10.1029/2002GL016724; 2003), where Greg Anderson and Chen Ji describe their modelling of stress transfer during this earthquake sequence. They conclude that the magnitude 6.7 earthquake of 23 October transferred 30–50 kilopascals of stress to the hypocentre of the

second (the Denali fault mainshock) — thereby, as the authors put it, "encouraging the occurrence" of the mainshock. From other studies, it seems that levels of 10–20 kPa of stress transfer can trigger further seismicity.

Both of the earthquakes occurred along the Denali fault system, in a mountain range some 90 miles south of Fairbanks. This is a 'strike-slip' fault, where two tectonic plates are in transform movement past one another, akin to the San Andreas fault that so exercises Californians. For context, the Denali fault mainshock was about the same magnitude as the San Francisco earthquake of 1906, with surface offsets

across the fault of up to 9 metres.

Like the San Andreas, the Denali fault system is complex, with many regional faults. Anderson and Ji also calculate the combined effect on those faults of the events of 23 October and 3 November. The resulting picture is varied, with minor stress release or increase on most fault segments, but in a couple of cases a massive increase of more than 400 kPa. This is well over the commonly cited threshold for triggering further seismic activity — showing that the threshold is perhaps not a 'hard floor', and that these fault segments might prove especially lively subjects for monitoring.

Tim Lincoln

deterministic component, corresponding to a heteroclinic cycle of unstable equilibria. The concept of a heteroclinic cycle deserves to be much more widely known, because it provides a simple and natural explanation of a puzzling but common phenomenon — a form of ‘punctuated equilibrium’. There is now the prospect that this elegant mathematical concept will help to answer

a long-standing question in meteorology: the mechanism of regime change. ■

Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK. e-mail: ins@maths.warwick.ac.uk

1. Cromellin, D. *Nonlinear Dynamics of Atmospheric Regime Transitions*. Thesis, Univ. Utrecht (2003).
2. Cromellin, D. *J. Atmos. Sci.* **59**, 1533–1549 (2002).
3. Palmer, T. N. *J. Clim.* **12**, 575–591 (1999).

Neurobiology

Dopamine as chicken and egg

David Self

State-of-the-art technology has allowed pulses of the neurotransmitter dopamine to be measured on a subsecond timescale in the brains of rats. It seems that dopamine both precedes and follows the pursuit of rewards.

Most readers will be familiar with the procedure used by animal trainers to get their animals to perform tricks: follow each successful performance with a tasty snack. This type of learning illustrates the phenomenon of reinforcement, whereby new behavioural responses increase in frequency when they are linked to rewards. At the molecular level, the neurotransmitter dopamine is important in reinforcement; it is released from brain cells in response to natural rewards, including food and during sexual arousal¹, and most drugs of abuse increase brain dopamine levels². Moreover, dopamine release in forebrain regions such as the nucleus accumbens is enough to reinforce new behaviours: laboratory animals will approach and press a lever repeatedly when dopamine is infused into their nucleus accumbens after each lever press³. But what are the neurochemical signals that actually ‘trigger’ reward-seeking behaviour as a result of dopamine-mediated reinforcement? Obviously, these signals must occur before, and not after, the reward is received. Phillips and colleagues⁴ help to resolve this dilemma by showing on page 614 of this issue that dopamine elicits reward-seeking behaviour, as well as signalling when a reward is received.

The story and the dilemma begin with B. F. Skinner⁵, who believed that a behavioural response is not really a ‘response’ to anything, but instead represents a spontaneously emitted act (or operant). The process of reinforcement is thought to strengthen connections between nerve cells and to ‘stamp in’ the inherent cellular activity underlying the spontaneous act. From a neurobiological perspective, some nerve cells do emit spontaneous bursts of electrical activity, and dopamine can increase the frequency of cell bursting, but only when it is applied immediately after the burst⁶ (akin to a reward for behaviour). So the idea, according to skinnerians, is that motivation to repeat a reinforced behaviour is automatic or

habitual, and there is no requirement for preceding external events to trigger it.

But external events can sometimes be highly effective triggers of reward-seeking behaviour, a phenomenon known as priming⁷. Chocolate lovers whose cravings are strongly enhanced by tasting just a small morsel often experience this priming effect. The initial taste whets the appetite for more, explaining the brief shelf-life of an open chocolate-box. Environmental cues also exert strong priming effects when they become associated with rewards through the process of pavlovian learning. Pavlov’s dogs learned to associate the sound of a bell with food, and then salivated at the sound of the bell alone. Similarly, other environmental cues can become associated with rewards and elicit a reward-seeking response. For instance, cues such as the sight of a drug syringe or associated paraphernalia are powerful triggers of craving in drug addicts.

So, reward-seeking behaviour might not always be spontaneous, as Skinner proposed, but rather the response to a reward-predicting environmental cue. Schultz⁸ found that such cues acquire the ability to activate dopamine release in the brain as they become associated with rewards through pavlovian learning. At the same time, rewards lose that ability. It is as if the brain reserves dopamine release for the most important events, signalling the occurrence of unexpected, novel rewards when learning is important, but not when those rewards were already anticipated thanks to a predictive cue. Yet the brain spares no dopamine with drugs such as cocaine, which bypass these dopamine-conserving mechanisms and increase dopamine release regardless of whether or not the drug reward is anticipated. In essence, such drugs prolong the learning phase indefinitely, and the brain continues to perceive cocaine as a novel reward despite years of use — perhaps explaining why the habit is difficult to break.

Reward-predicting environmental cues, then, can stimulate dopamine release — but it has been difficult to prove whether such cues actually ‘use’ dopamine to trigger reward-seeking behaviour. Definitive proof would require accurate subsecond measurements of dopamine release to determine whether dopamine pulses precede the initiation of reward-seeking behaviour. If they do, they could represent the neurochemical cause for the motivation to pursue reward. If they occur after initiation, they could simply reflect the role of dopamine in performing the behaviour.

To find out, Phillips *et al.*⁴ measured dopamine levels by implanting a carbon-fibre electrode into the nucleus accumbens of rats that had been trained to press a lever to receive a cocaine injection. Using state-of-the-art electrochemical technology called

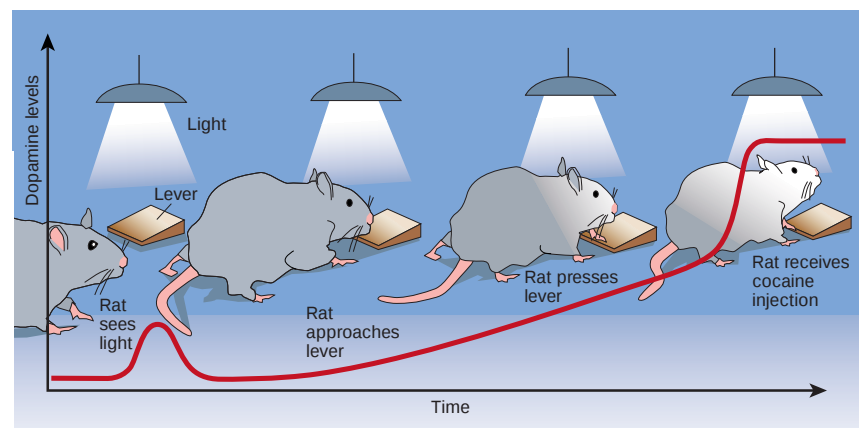


Figure 1 The neurotransmitter dopamine both motivates reward-seeking behaviour and signals the receipt of a reward. Phillips *et al.*⁴ studied subsecond changes in dopamine levels in the nucleus accumbens region of the brain in rats. The animals had been trained to press a lever to receive an injection of cocaine, and (in some experiments) had learnt to associate a flash of light and a tone with the drug. The authors detected a brief pulse of dopamine in response to the audiovisual cue; dopamine levels then continued to rise as the rats approached and pressed the lever, and (as found previously) peaked just after the animals received a cocaine injection — the ‘reward’.

fast-scan cyclic voltammetry, the authors could measure 100-millisecond pulses of dopamine — in the past, only minute-to-minute changes had been studied.

Phillips *et al.* found that a brief dopamine pulse was released just a few seconds before the animals became interested in approaching and pressing a lever that delivered cocaine injections. (This pulse is thought to be triggered by the context of the drug-associated test chamber, a pavlovian association.) Dopamine levels continued to increase during the final approach to the lever, and peaked just a few seconds after the lever press, while the cocaine injection and an audiovisual cue were delivered simultaneously. In animals that had learned to associate the audiovisual cue with cocaine, the cue itself caused a rapid increase in dopamine levels for several seconds (Fig. 1), even when it was presented in the absence of lever pressing or cocaine injections. This did not occur in animals that had never learned to associate the cue with cocaine. Finally, electrical stimulation of dopamine-producing neurons triggered a drug-seeking response — rats approached and pressed the lever.

These findings suggest that naturally evoked dopamine pulses are involved in both triggering and pursuit of reward-seeking behaviour, and that environmental cues (here, the test chamber or the audiovisual cue) use this mechanism to prime such behaviour. Previous studies detected prolonged increases in dopamine levels that lasted for several minutes following a drug reward. But drug-seeking behaviour ceases during these large, post-injection dopamine increases. Moreover, such studies generated

mixed, and even contradictory, findings regarding events that happen before the drug reward is received. Phillips *et al.*⁴ suggest that subsecond dopamine pulses are released only when drug-induced dopamine levels fall below a certain threshold. Similarly, they propose that environmental cues trigger reward seeking only when this subsecond regulation of dopamine pulses is possible.

It seems that dopamine represents both the chicken and the egg in the events underlying behavioural reinforcement. Dopamine acts as a reward for behaviour that precedes its release, and subsequently it triggers pursuit of the same reward after its release. As a rat chases its tail, so drug addicts may suffer a similar vicious circle of priming and reward controlled by these dopamine signals. Therapies aimed at preventing one or both of these signals could be effective treatments for addiction. ■

David Self is in the Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9070, USA.

e-mail: david.self@utsouthwestern.edu

1. Everitt, B. J. *Neurosci. Biobehav. Rev.* **14**, 217–232 (1990).
2. Di Chiara, G. & Imperato, A. *Proc. Natl Acad. Sci. USA* **85**, 5274–5278 (1988).
3. Dworkin, S. I., Goeters, N. E. & Smith, J. E. in *NIDA Research Monograph Vol. 67* (ed. Harris, L. S.) 242–248 (NIDA, Washington, 1986).
4. Phillips, P. E. M., Stuber, G. D., Heien, M. L. A. V., Wightman, R. M. & Carelli, R. M. *Nature* **422**, 614–618 (2003).
5. Skinner, B. F. *The Behavior of Organisms* (Appleton-Century-Crofts, New York, 1938).
6. Stein, L., Xue, B. G. & Belluzzi, J. D. *J. Exp. Anal. Behav.* **61**, 155–168 (1994).
7. Wise, R. A. in *The Neuropharmacological Basis of Reward* (eds Lieberman, J. M. & Cooper, S. J.) 377–424 (Oxford Univ. Press, New York, 1989).
8. Schultz, W. *J. Neurophysiol.* **80**, 1–27 (1998).

the one developed by Kominis *et al.*¹, detecting an external magnetic field can be thought of as a three-stage process (Fig. 1) — although, in fact, all three stages may occur simultaneously. First, the atoms are optically pumped and adopt the polarization of the pumping laser beam. Second, the magnetic field to be measured causes the atomic polarization to evolve: in the simplest case, the atomic magnetic moment precesses around the direction of the magnetic field. Finally, the evolution of the polarization is probed using laser light — either from the same laser beam that produced the pumping, or from a separate source (as it is in Kominis and colleagues' set-up).

According to the laws of quantum mechanics (namely, the uncertainty principle), for a given time-duration of the measurement, the intrinsic sensitivity to magnetic fields is inversely proportional to a product of three quantities. These are the magnetic moment of the atoms, the square root of the number of atoms involved in the measurement and the square root of the spin-relaxation time — that is, the length of time for which polarized atoms maintain their polarization in the absence of pumping. So, to improve the sensitivity beyond levels already achieved in other devices (which, for a measurement lasting one second, is at best about 1 femtotesla — $1 \text{ fT} = 10^{-15} \text{ T}$), the number of atoms in the system and their spin-relaxation time should be maximized.

There are several mechanisms that limit spin-relaxation time, one of the most important being depolarization caused by collisions with the cell walls that enclose the atomic vapour. The best way to get long spin-relaxation times has been to use vapour cells whose inner walls are coated with paraffin. An atom can bounce many thousands of times against a paraffin-coated wall before becoming depolarized; relaxation times are of the order of a second for a cell that is 10 cm in diameter. The atom density inside such cells, however, is usually relatively low, fewer than 10^{10} atoms per cubic centimetre: at higher densities the spin-relaxation time decreases because of spin-exchange collisions between the alkali atoms.

In a spin-exchange collision, the electron spins of the colliding atoms rotate with respect to their combined spin, which is conserved in the collision. All alkali atoms have non-zero nuclear spin, I , and their ground states are split into two 'hyperfine-structure' components, characterized by the total angular momentum $F = I \pm 1/2$. The direction of magnetic precession, determined by the relative orientation of the electron spin with respect to the total angular momentum, is opposite for the two hyperfine states. Thus, in the presence of a magnetic field, the spin-exchange collisions that randomly transfer atoms between these states normally lead to spin-relaxation, as atomic spins

Atomic physics

A new spin on magnetometry

Dmitry Budker

The capability to measure small, localized magnetic fields is valuable in biology as well as physics. A new device, based on spin-polarized alkali atoms, achieves better sensitivity and resolution than before.

On page 596 of this issue, I. K. Kominis and colleagues¹ report on a new type of atomic magnetometer that achieves unparalleled sensitivity to small magnetic fields as well as millimetre-scale spatial resolution. This device may open up exciting frontiers in diverse areas of science, ranging from biomagnetic imaging to testing fundamental symmetries of nature.

The most sensitive magnetometers of this type use vapours of alkali atoms, such as potassium, that have a single unpaired electron. When the atoms are subject to a beam of resonant, polarized light (such as from a laser), these unpaired electrons are excited to a higher energy level, but

subsequently return to their ground state by spontaneously emitting the energy or losing it through collisions. This 'pumping' process transfers the polarization of the incoming light to the atoms, affecting the direction of their electronic spins. But imparting polarization to the atoms like this, in turn, modifies the optical properties of the atomic medium; for example, it changes the way it absorbs laser light. The sensitivity of such a system to an external magnetic field comes from the torque that the field exerts on the magnetic moments of the atoms, which are themselves proportional to the electron spin.

In 'all-optical' magnetometers, such as

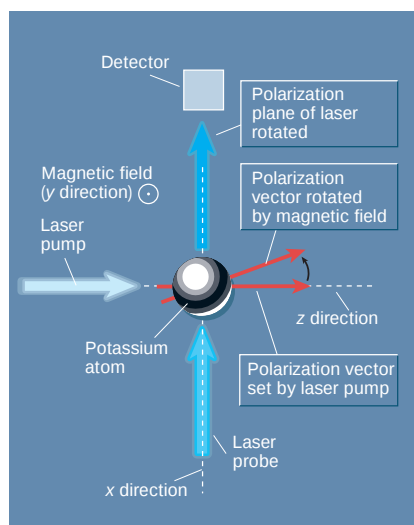


Figure 1 An 'all-optical' magnetometer. Potassium atoms are optically pumped by a laser beam, taking on its polarization. The presence of a magnetic field (in the y direction, out of the plane of the page) deflects the atoms' polarization vector. A second laser beam — the 'probe' — picks up this deflection, as its own polarization plane is rotated. Detecting this effect on the polarization of the laser probe gives a sensitive measure of the magnetic field strength. In the device developed by Kominis *et al.*¹, sensitivity to fields below 10^{-15} T has already been achieved, and 10^{-18} T could be reached.

acquire random angles with respect to each other.

Kominis *et al.*¹ have overcome the limitation due to spin exchange, and maintain an alkali-atom density in the vapour cell that is about 10,000 times higher than before — without any substantial degradation of the relaxation time. The idea (which goes back to the work of William Happer and colleagues in the 1970s) is that if the rate of spin-exchange collisions significantly exceeds that of the magnetic precession, the spin-exchange relaxation is effectively turned off. In this rapid spin-exchange regime, we may no longer speak of magnetic precession of atoms in an individual hyperfine state. Instead, each atom experiences an average precession in the same direction as would a free atom in the $F = I \pm 1/2$ state. This is because atoms, while being redistributed among the sublevels, spend more time in this state, which has a higher statistical weight and higher electron-spin polarization. As a result, in a weak external magnetic field, the average angular momentum of the atomic vapour precesses without spin-exchange relaxation (although at a slightly slower rate than free atoms).

To reduce spin-relaxation on the cell walls, instead of using an anti-relaxation wall coating Kominis *et al.* add a dense helium buffer gas (at a pressure of several atmospheres) to the alkali-atom vapour cell. The

presence of the buffer gas slows down the diffusion of the alkali atoms. This, combined with spatially resolved optical detection, means that localized magnetic-field measurements can be made.

This magnetometer seems ideal for biomagnetic imaging, particularly for recording electromagnetic activity in the brain. Optical-pumping magnetometers have been used successfully in the past to map the activity of the heart, which produces fields of the order of 10^{-10} T (see, for example, ref. 2). But brain imaging is more challenging because the magnetic fields in the brain are at least a thousand times weaker. The new device can detect a magnetic field as small as 0.5 fT, pinpointed to within 2 mm. And Kominis *et al.* claim that both the sensitivity and resolution could be improved with further development of the magnetometer — the magnetic-field sensitivity could even reach 10^{-18} T.

It is also interesting to note a related development: Kornack and Romalis³ (co-authors of Kominis *et al.*¹) have demonstrated a more unusual 'magnetometer' — designed to be completely insensitive to magnetic fields. In the vapour cell of this device, the buffer gas is another isotope of helium, ^3He , which becomes spin-polarized through spin

exchange with the optically pumped potassium atoms. The polarized helium picks up variations in the external magnetic field and automatically compensates for their effect on the potassium atoms, through the effective magnetic field that arises from spin-exchange interactions between the two. Instead of measuring a magnetic field, the aim here is to be sensitive to hypothetical interactions that might arise from any failure (or violation) of the fundamental symmetries of physics, such as Lorentz invariance, which holds that the laws of physics are the same for all observers independent of their positions and uniform motion in space.

This and the work of Kominis *et al.*¹ continue a productive tradition in atomic physics of synergy between fundamental and applied science.

Dmitry Budker is in the Department of Physics, University of California at Berkeley, Berkeley, California 94720-7300, USA.

e-mail: budker@socrates.berkeley.edu

1. Kominis, I. K., Kornack, T. W., Allred, J. C. & Romalis, M. V. *Nature* **422**, 596–599 (2003).
2. Bison, G., Wynands, R. & Weis, A. *Appl. Phys. B* doi:10.1007/s00340-003-1120-z (2003).
3. Kornack, T. W. & Romalis, M. V. *Phys. Rev. Lett.* **89**, 253002 (2002).

Nitrogen cycle

Solution to a marine mystery

Allan H. Devol

The anammox reaction, a microbial process that was first observed in waste-water treatment plants, looks as if it may be a key player in the nitrogen cycle in certain parts of the oceans.

In the oceans, denitrification is the process by which nitrate (primarily) is reduced to N_2 — inert dinitrogen gas. It occurs when certain bacteria decompose organic matter in environments where oxygen concentrations are vanishingly low, and is how 'fixed' nitrogen is converted into a form that cannot be used by most marine plants. This was believed to be the only mechanism of N_2 production in the oceans — and so by far the largest marine sink for fixed nitrogen.

On pages 606 and 608 of this issue, however, Dalsgaard *et al.*¹ and Kuypers *et al.*² show that large-scale conversion of fixed nitrogen to N_2 is probably occurring through another route. They find that N_2 can be produced by the anaerobic oxidation of ammonia in the oceanic water column, and that this 'anammox reaction' may be common in natural marine environments. Although there was evidence that the anammox reaction occurs in marine sediments³, the new findings enlarge the picture and could significantly alter our understanding of nitrogen cycling in the ocean. Denitrification is balanced by nitrogen fixation, which

is carried out in surface waters by highly specialized organisms that can reduce N_2 for incorporation into organic tissue. The balance between the two processes is such that, over vast areas, plant productivity in the oceans is limited by the amount of fixed nitrogen⁴. So the details of denitrification and nitrogen fixation are very important.

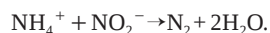
In the upper, sunlit layer of the oceans, photosynthetic organisms produce organic matter in the proportion carbon:nitrogen:phosphorus of about 106:16:1. This is the Redfield ratio. When this organic matter sinks out of the sunlit layer and is decomposed, carbon dioxide, ammonia and phosphate increase in the same ratio. In oxygenated marine environments, the regenerated ammonia is oxidized to nitrate. But in oxygen-deficient environments, decomposition continues via denitrification, and nitrate concentrations begin to decrease while those of phosphate continue to increase. Oceanographers can estimate the amount of denitrification that has taken place from the difference between the amount of regenerated nitrate predicted

from the measured phosphate concentration and the nitrate actually measured⁵. This calculation yields the amount of nitrate reduced to N₂ during denitrification, and this decomposition should result in the 'remineralization' of ammonia — which, in the absence of oxygen, it was thought should remain as ammonia.

Many years ago, however, Richards⁶ pointed out that this ammonia could not be accounted for. He proposed that it might somehow be oxidized anaerobically to N₂ gas by nitrite or nitrate, either inorganically or by some unknown microbe. Geochemical evidence that this reaction does indeed occur is strikingly clear in oxygen-deficient environments such as the Black Sea⁷. In 1999, an organism that could carry out this 'anammox' reaction was isolated from a bioreactor system designed to remove ammonia from waste water, and was identified as a relative of a group of bacteria called the planctomycetes⁸. But it was thought unlikely that these organisms were either common or of much importance in the environment⁹; and despite the geochemical evidence, and the existence of a suspect organism, this nitrogen-cycle pathway remained a mystery.

Using an elegant bit of sleuthing, the two groups^{1,2} have now found the smoking gun that shows that the anammox reaction can, indeed, proceed under environmental conditions in the oceans, and that the likely perpetrators are closely related to the organisms isolated from bioreactors. Both groups worked in suboxic areas where the geochemical distributions of nitrogen species suggested that the anammox process was occurring (see Fig. 1 of each paper on pages 606 and 609).

Dalsgaard *et al.*¹ used isotope tracers of nitrogen to show that when radiolabelled ammonia (¹⁵NH₃) was added to nitrate-containing water-column samples from the oxygen-depleted zone of an inlet on the west coast of Costa Rica, the ¹⁵N was incorporated into N₂ when incubated in quasi-natural conditions. Furthermore, when both ¹⁵NH₃ and isotopically labelled nitrate (¹⁵NO₃⁻) were added, the relative yield of ²⁸N₂, ²⁹N₂ and ³⁰N₂ showed that the anammox reaction followed the form



This conversion of ammonium and nitrite to N₂ and water is the same as the anammox reaction seen in the waste-water bioreactor.

Similarly, Kuypers *et al.*² show that ¹⁵NH₃, added to samples from the oxygen-depleted zone in the Black Sea, was converted to N₂ in the suboxic zone, substantiating the geochemical evidence⁷ that the anammox reaction is occurring there. To see if the anammox organism in the Black Sea is

similar to the well-characterized planctomycete isolates, Kuypers *et al.* looked for the unique signature of strange lipids known as ladderanes. These lipids surround the anammoxosome, the unique, organelle-like structure within which the anammox reaction takes place. They found that the distribution of ladderanes in the water column closely mirrored the distribution of the anammox reaction as indicated by their nitrogen-isotope experiments. Furthermore, molecular DNA analyses showed that organisms closely related to the bioreactor planctomycetes were present and common at the relevant depths. Taken together, these studies^{1,2} present pretty compelling evidence that the anammox reaction takes place in the water column in natural marine environments, and that it is probably carried out by planctomycetes similar to those isolated from bioreactors.

So, just how important is this reaction in the oceans? Conditions favourable for denitrification occur primarily in the oxygen-deficient zones of water columns off the west coasts of Central and South America and India, and also in marine sediments, mainly on continental margins. The total amount is more or less equally partitioned between the water column and sediments⁴. Current estimates of water-column denitrification are based mostly on budget studies of nitrogen and phosphorus, and the Redfield ratio as outlined above. If anammox is responsible for the oxidation of regenerated ammonia, then 28% of the N₂ production would be attributed to this reaction. If proteins are the preferential substrate for denitrification¹⁰, this percentage increases to 48%. Perhaps more importantly, sulphate reduction deeper in marine sediments produces large quantities of ammonia, which diffuse up into the denitrification zone where anammox could, again, rival denitrification as an N₂-producing process.

All in all, it looks possible that anammox may account for between 30% and 50% of the N₂ production in the oceans. If this is so, we will have to revise our ideas about the mechanisms of marine denitrification and how we think about the marine nitrogen budget. ■

Allan H. Devol is in the School of Oceanography, University of Washington, Box 357940, Seattle, Washington 98195-7940, USA.

e-mail: devol@u.washington.edu

- Dalsgaard, T., Canfield, D. E., Peterson, J., Thamdrup, B. & Acuña-González, J. *Nature* **422**, 606–608 (2003).
- Kuypers, M. M. M. *et al.* *Nature* **422**, 608–611 (2003).
- Thamdrup, B. & Dalsgaard, T. *Appl. Environ. Microbiol.* **68**, 1312–1318 (2002).
- Codispoti, L. A. *Nature* **387**, 237–238 (1997).
- Gruber, N. & Sarmiento, J. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
- Richards, F. A. in *Chemical Oceanography* (eds Riley, J. P. & Skirrow, G.) 611–645 (Academic, London, 1965).
- Murray, J. W., Codispoti, L. A. & Fredrich, G. W. in *Aquatic Chemistry* (eds Huang, C. P., O'Melia, C. R. & Morgan, J. J.) 157–176 (Am. Chem. Soc., Washington DC, 1995).
- Strous, M. *et al.* *Nature* **400**, 446–449 (1999).
- Zehr, J. P. & Ward, B. B. *Appl. Environ. Microbiol.* **68**, 1015–1024 (2002).
- Van Mooy, B. A. S., Keil, R. G. & Devol, A. H. *Geochim. Cosmochim. Acta* **66**, 457–465 (2002).



100 YEARS AGO

Prof. H. H. Turner, Savilian professor of astronomy in the University of Oxford, contributes to the *Fortnightly Review* for April a reply to Dr. Wallace's article on "Man's Place in the Universe" which was published in the same review last month. Dr. Wallace suggested that the universe is limited in extent; that it has a definite centre at which the solar system is, and has been situated for millions of years; and that by reason of its position the earth has had an opportunity to develop humanity, and probably this opportunity has been nowhere else in the universe. Prof. Turner shows that the limitation in the universe is not proved; that there is no true centre of the universe, even if limited, and even if there were the solar system could not occupy it for long, on account of the sun's proper motion; he also shows that there is no reason whatever why life should not be developed in any part of the interior of even a limited universe. From *Nature* 9 April 1903.

50 YEARS AGO

It is now widely recognized that, in a society with an ageing population, it is desirable and even essential that the elderly should remain longer at work... the first and most essential step, Mr. Hopkins argues, is to eliminate from the national insurance system the commitment to provide a retirement pension at a definite age laid down years, or even decades, in advance. Mr. Hopkins recognizes that the presentation of the scheme to the public in the past has been such as to cause false impressions and to create expectations which may lead to resentment when government is compelled, as he believes is inevitable, to raise the pension age. Nevertheless, he believes it is necessary that the principle should be publicly established that the age at which retirement pension may be claimed is variable, and will probably be raised as time goes on. As he admits, a government which made such an announcement would obtain no immediate credit, and would probably incur some immediate odium. It would, however, earn the gratitude of its successors by establishing the conditions in which pension arrangements could eventually be adjusted to changes in the demographic and economic circumstances of the country. No political economic party has, however, yet given any signs of the courage demanded in taking such a step. From *Nature* 11 April 1953.

Physical chemistry

Salts leave fingerprints in acid

J. Phys. Chem. B **107**, 2875–2878 (2003)

Adding salt to a solution can be a perplexing business. First, it can change the pH, but even more puzzlingly, the effect depends on the chemical identity of the ions — not just on their charge, but on whether they are bromide, chloride or whatever.

Such effects have been reported before, but Mathias Boström *et al.* have subjected them to close scrutiny. They find, for example, that the pH of two otherwise identical solutions, measured using a standard glass-electrode technique, can differ by up to 0.5 for identical concentrations of potassium bromide and sodium perchlorate.

This specificity of an added salt recalls the well-known Hofmeister effect: the tendency of different ions to precipitate proteins from solution. It is a mystery thought to be bound up with the effect of ions on water's structure, and has implications for the way particular salts modify the cell environment.

Boström *et al.* suggest an explanation: that the pH effect arises because ions close to the interface of the water and the glass electrode are not distributed evenly, which in turn affects ion-specific electrostatic dispersion forces at this interface. These forces can alter the way hydronium (H_3O^+) ions interact with the glass surface — precisely the kind of interaction picked up by the glass-electrode technique. **Philip Ball**

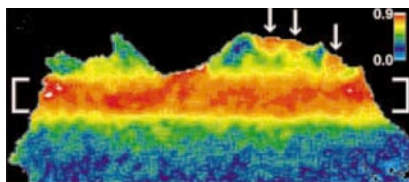
Cell biology

Fast-track for actin

Science **300**, 142–145 (2003)

Cells move or change shape by making polymers of actin protein at their leading edge — the resulting actin filaments then push on the cell membrane, forcing it outwards. It is commonly thought that the actin monomers required for polymerization reach the leading edge by the random, rather slow process of diffusion. Daniel Zicha and colleagues now show that this is not always the case.

Using a microscopic technique called 'fluorescence localization after photobleaching', Zicha *et al.* tracked actin monomers moving over large distances within live cells. Combined images of cells containing actin molecules labelled with two different fluorescent proteins, obtained after photobleaching one of the fluorescent labels in a small area of the cell (white brackets in picture), allow the fate of the bleached actin to be followed. This shows a strikingly swift movement of actin to the rapidly protruding edges of the cell (white arrows) — far faster than could be explained by simple diffusion.



Leading edge: images of actin in a cell on the move.

The authors suggest that a hydrodynamic flow created by the cellular shape changes themselves promotes the fast actin movement they have seen. **Mirella Buccì**

Human evolution

Being manipulative

J. Hum. Evol. **44**, 225–254 (2003)

Lucy's hands were as deft as ours, according to a new analysis of fossil hand-bones of *Australopithecus afarensis* — the species to which Lucy belongs. The hands of *A. afarensis* have previously been described as chimp-like. But based on their study of several different skeletons found at one site in Ethiopia, David M. Alba and colleagues conclude that the hand proportions were "fully human". They claim that the finding throws into doubt the idea that human-like hands evolved as an adaptation for tool-making. There is no evidence that *A. afarensis*, which lived in East Africa about 3.5 million years ago, made stone tools: the first tools only appear in the archaeological record about a million years later.

Australopithecus afarensis was bipedal. Released from the demands of locomotion, the hands may have been free to respond to pre-existing evolutionary pressures, perhaps for grooming and feeding, and were later co-opted for making tools. Humans have shorter fingers and so relatively longer thumbs than our ape cousins. This allows the tips of all five digits to be brought together, making our hands better at manipulating objects. **John Whitfield**

Immunology

A numbers game

Immunity **18**, 343–354 (2003)

Our cells are covered with tiny protein fragments that are bound to class I molecules of the major histocompatibility complex (MHC). They tell the immune system about the proteins being produced in the cell — whether it is infected by a virus, for instance. It takes fewer than ten MHC class I molecules bearing their viral fragment load to launch an immune response.

So far, so efficient — but what does it take to display ten such complexes on the cell surface? Michael F. Princiotta and colleagues have studied the economics of the process, and tell a tale of staggering numbers and apparently enormous waste.

They find that cells maintain a content

of around 2.6 billion proteins, using about 6 million ribosome particles to make them, at a rate of 4 million proteins per minute. More than 30% are immediately degraded, at a rate of 2 million proteins per minute. Amidst this wholesale demolition, protein fragments are produced for presentation by MHC class I molecules. But only 1 in 40 degraded proteins results in a suitable fragment, and only 1 in 50 fragments makes it to the endoplasmic reticulum, where it can meet with an MHC class I molecule. So 2,000 proteins need to be degraded for a single fragment to be displayed.

Wasteful? Perhaps. But, as the authors point out, "ultimately, selection pressure in evolution is exerted by other similarly inefficient organisms and not arbitrary standards of efficiency conjured by human intelligence" **Marie-Thérèse Heemels**

Nanotechnology

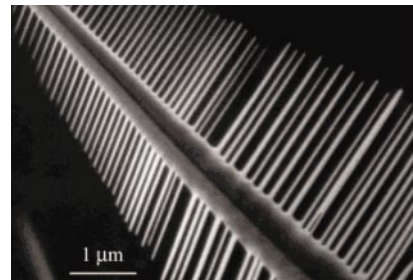
Dendritic lasers

J. Am. Chem. Soc., advance online publication doi:10.1021/ja034327m (2003)

Nanowires that act like tiny lasers have been grown in a parallel, orderly array like the teeth of a comb, producing a bank of miniature light sources that might be useful for photonic technology and nanoscale electromechanical devices. Significantly, they demand none of the complex fabrication procedures used in conventional device manufacture for microelectronics and optoelectronics.

The nanowire 'combs' made by Haoquan Yan *et al.* (see picture) are the products of spontaneous self-organization. The researchers grew needles of zinc oxide, a light-emitting semiconductor, from zinc powder heated with oxygen gas. The needles, about a micrometre or so wide, are decorated with dendritic side-branches regularly spaced about 0.1–2 μm apart, each branch a crystalline rod of zinc oxide about 10–300 nm wide.

When pumped with laser light, these arrays of wires produced emission in the ultraviolet that had the narrow wavelength band characteristic of lasing. For arrays with larger rods and larger spaces, the spots of light from each nanowire tip could be resolved individually. **Philip Ball**



Fine-tooth comb: zinc-oxide nanowires.

Floating gold in cryogenic oxygen

This fluid can be used to create remarkable magnetic levitation and patterning effects.

In magnetic levitation, a strong and spatially varying magnetic field exerts an upward force on a body that is sufficient to counteract its weight due to gravity. Here we show that this effect can be enhanced by immersing the body in cold oxygen gas, which provides a further strong and adjustable buoyancy force that allows a wide range of materials to be levitated in an open, unpressurized vessel. The buoyancy of magnetized liquid oxygen is sufficient to float even gold and platinum, suggesting that this technique could find application in mineral separation. An interesting periodic pattern is created on the surface of this pure elemental paramagnetic fluid.

Magnetic levitation occurs when the force on a diamagnetic body, due to an inhomogeneous magnetic field, is strong enough to balance the body's weight due to Earth's gravity¹. The levitation force per unit volume of the body is given by $(\chi/\mu_0)BdB/dz$, where χ is the volume magnetic susceptibility, μ_0 is the permeability of free space, B is the magnetic field and dB/dz is the vertical field gradient. Immersing the body in a paramagnetic fluid, such as gaseous oxygen, enhances the levitation by providing buoyancy through the magneto-Archimedes effect. This technique has been used to levitate sodium chloride (density, 2.17 g ml⁻¹) in a closed vessel containing oxygen at a pressure of 60 atmospheres (6 megapascals) at room temperature².

We exploit two well-known laws of physics to levitate a wider range of diamagnetic materials, including crystals of quartz (2.65 g ml⁻¹), potassium bromide (2.75 g ml⁻¹) and diamond (3.51 g ml⁻¹), in an open vessel of cold oxygen gas at atmospheric pressure. According to Boyle's law, at a fixed pressure the gas density varies inversely with temperature; Curie's law states that the paramagnetic susceptibility is also inversely proportional to the temperature. Near the boiling point of liquid oxygen (90 K), the magnetic buoyancy of oxygen gas is therefore about ten times stronger than at room temperature. This enhanced levitation is shown in Fig. 1a (and see movie in supplementary information).

Liquid oxygen provides still greater buoyancy, which is sufficient to float, at relatively low magnetic-field strengths, the densest diamagnetic objects and even weakly paramagnetic objects. Figure 1b shows a silicon crystal, a gallium arsenide crystal, a £1 coin, a piece of lead and a gold coin, all floating in the liquid oxygen. Each object floats at the point at which the local

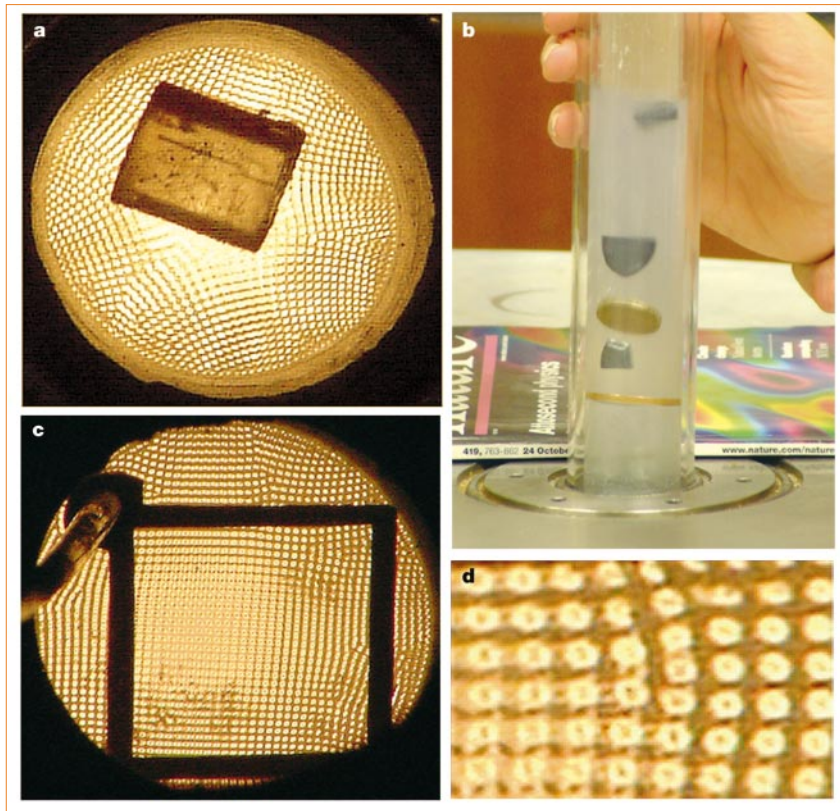


Figure 1 Levitation, flotation and surface-instability patterns obtained by using cryogenic oxygen and high magnetic fields. **a**, Potassium chloride crystal levitating near the top of the magnet in an atmosphere of cold oxygen gas at about 200 K; $BdB/dz = 1,400 \text{ T}^2 \text{ m}^{-1}$. In the background is a liquid-oxygen membrane located at the centre of the magnet bore, where $B = 17 \text{ T}$. The liquid's surface is decorated by an instability pattern. **b**, Dense objects floating in liquid oxygen in a glass Dewar vessel positioned just above the top of the magnet bore. Top to bottom: a silicon crystal, a gallium arsenide crystal, a £1 coin, a piece of lead and a gold coin. The field varies from 0.7 T (silicon) to 2 T (gold). **c**, A rectangular copper frame encloses a square-lattice instability pattern on the surface of liquid oxygen at 17 T. Several edge dislocations can be seen to the right. **d**, Detail showing an edge dislocation in the pattern.

magneto-Archimedes force balances its weight, so its levitation position can be adjusted by changing the magnetic-field strength. This feature could be of use in mineral-separation technology.

We also noted the formation of a regular pattern of peaks on the surface of slowly boiling liquid oxygen; this pattern is held as a membrane at the centre of the magnet bore (Fig. 1a, c, d). The peaks arise from a combination of magnetic and surface-energy effects and have been observed before in synthetic ferrofluids³, but their appearance in a pure elemental paramagnetic liquid merits further investigation.

At high magnetic-field strengths ($B > 13$ tesla), a simple square-lattice structure, decorated by edge dislocations, can be formed inside a small rectangular frame. In weaker fields, hexagonal close-packed crystallites are evident. When the

heat input to the liquid oxygen is sufficiently great, the fluid exhibits convection and turbulence, which causes complicated motion of the microcrystallites and dislocations (see supplementary information). This could provide an experimental test bed for models of crystal formation and dynamics.

A. T. Catherall, L. Eaves, P. J. King, S. R. Booth

School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK
e-mail: laurence.eaves@nottingham.ac.uk

- Geim, A. K., Simon, M. D., Boamfa, M. I. & Heflinger, L. O. *Nature* **400**, 323–324 (1999).
- Ikezoe, Y., Hirota, N., Nakagawa, J. & Kitazawa, K. *Nature* **393**, 749–750 (1998).
- Cowley, M. D. & Rosensweig, R. E. *J. Fluid Mech.* **30**, 671–688 (1967).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Avian behaviour

Altruism and infidelity among warblers

In cooperatively breeding vertebrates, indirect fitness benefits^{1,2} are maximized by subordinates who choose to help their own closely related kin after accurately assessing their relatedness to the group's offspring^{3–5}. Here we show that in the cooperatively breeding Seychelles warbler (*Acrocephalus sechellensis*), female subordinates help to raise new nestlings by providing them with food only when the offspring are being raised by parents who also fed the subordinates themselves when they were young³. These helper females use the continued presence of the primary female, rather than of the primary male, as their provisioning cue — presumably because female infidelity is rife in this species⁶, making their relatedness to the father less reliable.

In *A. sechellensis*, female offspring normally become subordinate within their natal group^{7,8} and often help to raise dependent young⁷. We studied the provisioning behaviour (feeds given per hour) of 21 colour-ringed female subordinates. Reproductive sharing and high levels of extra-pair paternity occur in this species⁶; we therefore assessed parentage and genetic relatedness by using microsatellite DNA markers⁸. Subordinate parents (that is, subordinates who had also produced a nestling in the helped nest) always provisioned and were excluded from subsequent analysis ($n = 8$).

In a general linear model analysis, the provisioning behaviour of non-parent subordinates was influenced positively by subordinate–nestling relatedness ($F_{1,11} = 4.82$, $P = 0.05$, $R^2 = 0.31$), whereas subordinate age, sex, number of nestlings, number of subordinates (or feeding adults) and territory quality had no significant effect. A comparison of relatedness⁹ to the nestlings of subordinates who helped with those that did not (Fig. 1a) indicates that helpers were significantly related to nestling(s), whereas non-helpers were not (mean relatedness, 0.27 versus -0.05 , respectively).

The continued presence of the primary female, but not of the primary male, was a reliable cue of subordinate–nestling relatedness; subordinate–primary female relatedness predicted subordinate–nestling relatedness ($F_{1,38} = 27.37$, $P < 0.001$), but subordinate–primary male relatedness did not ($F_{2,37} = 2.11$, $P = 0.16$). Subordinates provisioned significantly more often when their putative mother was still present, but provisioning was unaffected by the presence or absence of the putative father (Fig. 1b).

Subordinates may determine when to provision by directly assessing their relatedness to either the nestling(s) or the primary

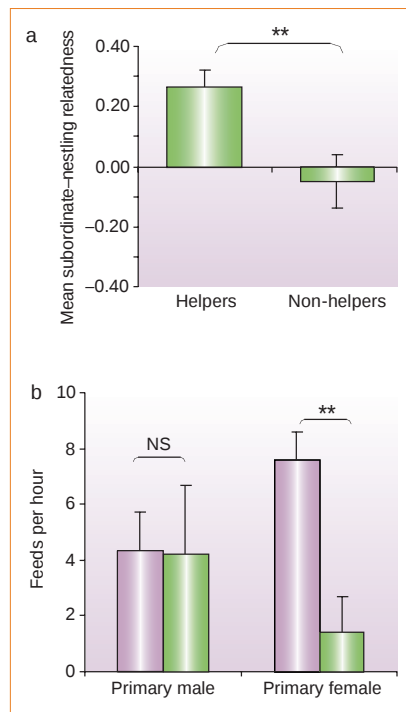


Figure 1 Provisioning by subordinate female, non-parent Seychelles warblers (*Acrocephalus sechellensis*). **a**, Provisioning in relation to the pairwise relatedness between subordinates and nestlings (subordinate helper–nestling relatedness versus non-related: 0.27 ± 0.16 versus 0 , one-sample t -test, $t_7 = 4.64$, $P < 0.002$; helpers versus non-helpers: 0.27 ± 0.16 , $n = 8$ versus -0.05 ± 0.20 , $n = 5$; $t_{11} = 3.13$, $P = 0.01$). **b**, Provisioning in relation to the presence (purple bars) or absence (green bars) of the subordinate's putative parents (presence versus absence of putative mother: 7.57 ± 2.44 , $n = 6$ versus 1.43 ± 3.60 , $n = 7$; t -test, $t_{11} = 3.71$, $P = 0.003$; presence versus absence of putative father: 4.30 ± 4.89 , $n = 4$ versus 4.24 ± 4.29 , $n = 8$; $t_{11} = 0.02$, $P = 0.98$). Error bars, means $\pm$ s.e.; asterisks indicate statistical significance.

female. However, the continued presence of the primary female explained provisioning more reliably ($F_{1,11} = 13.72$, $P = 0.003$, $R^2 = 0.55$) than genetic estimates of subordinate–nestling or subordinate–primary female relatedness ($F_{1,11} = 2.14$, $P = 0.17$; $F_{1,11} = 0.06$, $P = 0.81$, respectively). These results confirm that the presence of the primary female who raised the subordinate is likely to be used as a cue to determine when to provision. Previous findings that helpers can distinguish between siblings and half-siblings³ may have been an artefact caused by helpers assisting only their mothers.

Our results show that female subordinates can use an indirect but reliable cue to assess their relatedness to nestlings, and that this assessment of kinship determines provisioning rates. In *A. sechellensis*, the amount of food brought to the nestling determines fledging success and first-year survival, and the presence of a helper increases the number of young who are fledged on a territory¹⁰. The preferential provisioning of related nestlings by female subordinates will therefore increase the

number of related offspring produced. In our study, the number of fledglings produced on a territory increased significantly (by 17%) for each subordinate present (after exclusion of direct parentage)⁸.

Both direct⁸ and indirect benefits may be important within this cooperative breeding system as, by using effective discrimination, subordinates are able to maximize the indirect benefit gained within a system that is driven primarily by direct benefits⁸. Our findings show that, in the presence of a high frequency of female infidelity, an associative learning mechanism has evolved to focus on the mother at the nest — the only sex to which the subordinates are reliably related.

David S. Richardson*†, Jan Komdeur‡, Terry Burke*

*Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

†Department of Animal Ecology, University of Lund, 223 62 Lund, Sweden

e-mail: david.richardson@zoekol.lu.se

‡Zoological Laboratory, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands, and Department of Zoology, University of Melbourne, Victoria 3010, Australia

1. Hamilton, W. D. J. *Theor. Biol.* **7**, 1–52 (1964).
2. Emlen, S. T. *Proc. Natl Acad. Sci. USA* **92**, 8092–8099 (1995).
3. Komdeur, J. *Proc. R. Soc. Lond. B* **256**, 47–52 (1994).
4. Russell, A. F. & Hatchwell, B. J. *Proc. R. Soc. Lond. B* **268**, 2169–2174 (2001).
5. Hatchwell, B. J., Ross, D. J., Fowle, M. K. & McGowan, A. *Proc. R. Soc. Lond. B* **268**, 885–890 (2001).
6. Richardson, D., Jury, F., Blaakmeer, K., Komdeur, J. & Burke, T. *Mol. Ecol.* **10**, 2263–2273 (2001).
7. Komdeur, J. *Proc. R. Soc. Lond. B* **263**, 661–666 (1996).
8. Richardson, D. S., Burke, T. & Komdeur, J. *Evolution* **56**, 2313–2321 (2002).
9. Goodnight, K. F. & Queller, D. C. *Mol. Ecol.* **8**, 1231–1234 (1999).
10. Komdeur, J. *Behav. Ecol. Sociobiol.* **34**, 175–186 (1994).

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Plant ecology

Tree-species competition and coexistence

How apparently similar plant species coexist is a puzzle. Kelly and Bowler¹ claim that environmental fluctuation promotes the coexistence of tree species by alternately favouring recruitment of common and rare congeners in a dry tropical forest. Here I argue that current knowledge of tropical-forest ecology does not support the authors' focus on congeneric competition, and show that their use of diameter distributions to date recruitment fluctuations may be misleading. It is therefore doubtful, at this stage, that recruitment patterns of the authors' congeneric pairs can be linked to the sort of competition dynamic that they envisage.

Kelly and Bowler's two-component lottery model assumes that seedlings of

congeners compete for the same set of establishment sites, the availability of which is little affected by other taxa. Empirical evidence from other tropical forests suggests, however, that this is unlikely. Although it is reasonable to assume that closely related taxa have similar requirements, most tropical trees seem to have broad, overlapping niches^{2–4} and so are likely to compete for establishment sites with a wide variety of taxa, and not just with close relatives. There is little reason, therefore, to argue that a 'see-saw' model of interaction between congeneric pairs, discounting diffuse competition, is an informative approach to the dynamics of a forest containing more than 200 tree species.

Kelly and Bowler present evidence of out-of-phase recruitment fluctuations in rare and common congeners. Recruitment fluctuations were inferred from proxy age distributions, dividing trunk diameters by species-specific mean recent growth rates to estimate individual trees' ages. A problem with this procedure stems from the assumption that diameter distributions are a reliable indicator of age distributions. The fragility of this assumption is well known to dendro-ecologists working in temperate forests^{5–8}, where growth-ring counts allow ages to be estimated accurately. The often-poor correspondence between diameter distribution and age structure^{5–9} probably arises because stand dynamics have subjected different age classes to different growth histories, as well as spreading a given age class across a wide range of diameters.

The implications for Kelly and Bowler's method are exemplified by data from a temperate forest⁹ showing that 'recruitment fluctuations' inferred from diameter distributions and age–diameter relationships do not necessarily coincide with those suggested by the actual age data (Fig. 1). Therefore, although tree-diameter distributions can reasonably be used to draw broad inferences about age structure (such as distinguishing

between all-aged and even-aged populations), attempting to date recruitment fluctuations by such a method is probably overly ambitious^{5–7}.

In view of the difficulties of reliably dating recruitment fluctuations, and the questionable basis of the model, the case for mediation of coexistence of congeners by environmental fluctuation is unconvincing. Although there is every reason to believe that environmental instability contributes to coexistence in general¹⁰, establishing even the potential for its special relevance to the regeneration of congeneric trees in tropical forests would require clear evidence that seedlings of coexisting congeners overlap far more in space and time than they do with those of more distantly related associates.

Chris Lusk

Departamento de Botánica, Universidad de Concepción, Casilla 160-C, Concepción, Chile
e-mail: clusk@udec.cl

1. Kelly, C. K. & Bowler, M. G. *Nature* **417**, 437–440 (2002).
2. Lieberman, M., Lieberman, D. & Peralta, R. J. *Trop. Ecol.* **11**, 161–178 (1995).
3. Hubbell, S. P. *et al. Science* **283**, 554–557 (1999).
4. Harms, K. E., Condit, R., Hubbell, S. P. & Foster, R. B. *J. Ecol.* **89**, 947–959 (2001).
5. Harper, J. L. *Population Biology of Plants* (Academic, London, 1977).
6. Ogden, J. J. *Biogeogr.* **8**, 405–420 (1981).
7. Lorimer, C. G. *Can. J. For. Res.* **15**, 200–213 (1985).
8. Stewart, G. H. & Rose, A. B. *Vegetatio* **87**, 101–114 (1990).
9. Lusk, C. & Ogden, J. J. *Ecol.* **80**, 379–390 (1992).
10. Chesson, P. L. & Warner, R. R. *Am. Nat.* **117**, 923–943 (1981).

Plant ecology

Coexistence of tropical tree species

For decades, ecologists have struggled to explain how so many tropical tree species can coexist. Kelly and Bowler¹ propose that differences in recruitment fluctuation and competitive abilities among closely related tree species could promote coexistence, and data from a tropical deciduous forest in western Mexico seem to confirm their predictions. We argue, however, that the tests of their model's predictions make fundamentally flawed assumptions about both size–age relationships in trees and the factors that influence population size structures. As such, their results are potentially misleading and lack the necessary rigour to 'reject all other theories of coexistence'.

Kelly and Bowler's model of species coexistence predicts that rare species should exhibit greater deviation in recruitment than common species. To test this, they compared differences between expected and observed age-class distributions for ecologically similar congeners. In essence, they predicted that abundant species should show fewer deviations in historical recruitment trends, and thus have 'smoother' age-class distributions.

Determining the age of tropical trees is

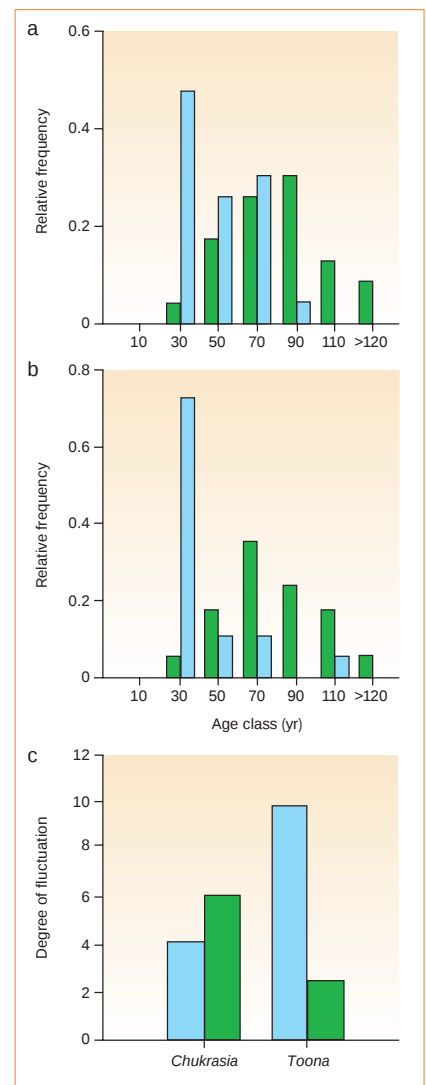


Figure 1 Comparisons of real (blue bars) and size-derived (green bars) age distributions for two tropical tree species that form annual growth rings. **a**, *Chukrasia tabularis*; **b**, *Toona ciliata*. **c**, Degree-of-fluctuation (d) values for real and size-derived age distributions. Data were obtained from canopy trees in a 50-hectare, permanent forest-dynamics plot at the Huai Kha Khaeng wildlife sanctuary, western Thailand. Real age distributions were obtained from tree-ring analysis. Estimated tree ages were calculated from cored trees by dividing stem diameter by a taxon-specific mean annual growth rate. Note that the estimated age distributions differ in both shape and scale from the real age distributions and that the values of d for the estimated distributions provide no indication of d for the real age distributions.

challenging because most such species do not form annual growth rings². Kelly and Bowler circumvent this problem by using average diameter-growth rates to convert tree size to age (that is, age = size/growth rate). This, however, makes the improbable assumption that average lifetime growth rates are the same for every tree in a species' population. Evidence from both temperate and tropical forests contradicts this assumption^{3,4}. Tree growth is indeterminate, and in natural forests the availability



Figure 1 Recruitment fluctuations as indicated by population age structure (green bars) and by use of stem diameter as a proxy for age (orange bars) in a population of *Weinmannia racemosa* ($n=216$) in an old-growth temperate forest in New Zealand⁹. Fluctuations are represented as proportional deviations (observed – expected/expected) from a negative exponential curve¹. Despite the highly significant overall relationship between age and diameter ($r=0.81$, $P<0.0001$), there is no correlation between 'real' and proxy fluctuations ($r=0.03$, $P=0.94$).

of resources required for growth varies in time and space. Trees of the same age may thus be small or large, depending on their individual growth history, making a species' size distribution an unreliable surrogate for its age distribution.

In Fig. 1 we present size and age data from two tropical tree species to illustrate the shortcomings of inferring historical recruitment fluctuations from age distributions derived from size distributions. Our aim is to demonstrate the potential errors involved in converting size- to age-class distributions and their influence on the degree-of-fluctuation statistic, d . Because d is, by design, sensitive to minor deviations from an expected distribution, it is not surprising that the values of d obtained from the size-derived age distribution in Fig. 1 bear no relationship to those from the real age distribution. Consequently, inference of abundance-related patterns in the magnitude of d obtained from size-derived age distributions is meaningless.

A more fundamental problem with Kelly and Bowler's analysis is the assumption that deviations from expected age-class distributions are solely due to temporal fluctuations in recruitment. Size-class distributions are complex functions of three factors: recruitment, growth and mortality. To interpret deviations in size- or age-class distribution simply as recruitment fluctuations is to assume that average growth and mortality rates are constant over time. Stochastic events such as wind storms, fires or climatic anomalies may alter growth and mortality rates for years or decades. Disentangling the effects of growth, mortality and recruitment on a continuous distribution of tree sizes is impossible without further information on past growing conditions — data that are rarely available for tropical forests.

Significant advances have been made in the study of tropical-forest dynamics at the scale of years, through large-scale forest-dynamics plots⁵, and millennia, through palaeoecological studies⁶. Neither has provided a satisfactory answer to the question of tree-species coexistence. Kelly and Bowler refocus the question on the largely unexplored middle ground — the scale of decades to centuries. However, lacking real tree ages, their theory remains untested.

Patrick J. Baker*, Jeremy S. Wilson†

**Institute of Pacific Islands Forestry, Pacific Southwest Research Station, USDA Forest Service, Hilo, Hawaii 96720, USA*

e-mail: patrickbaker@fs.fed.us

†Department of Forest Management, University of Maine, Orono, Maine 04469, USA

1. Kelly, C. K. & Bowler, M. G. *Nature* **417**, 437–440 (2002).
2. Bormann, F. H. & Berlyn, G. P. (eds) *Age and Growth Rates of Tropical Trees: New Directions for Research* (Yale Univ. Sch. Forestry Environ. Stud. Bull. 94, New Haven, Connecticut, 1981).
3. Richards, P. W. *The Tropical Rain Forest* (Cambridge Univ. Press, Cambridge, 1996).

4. Oliver, C. D. & Larson, B. C. *Forest Stand Dynamics* (Wiley, New York, 1996).
5. Condit, R. *Trends Ecol. Evol.* **10**, 18–22 (1995).
6. Bush, M. B. J. *Biogeogr.* **21**, 5–17 (1994).

Kelly and Bowler reply — Both comments concern our use¹ of the relationship between tree age and tree diameter. However, none of the points raised invalidates our original conclusion regarding storage dynamics. Lusk asks for independent evidence that competition is not diffuse, and argues that recruitment fluctuations cannot be dated reliably from fluctuations in size distribution. He highlights a key result that preceded our model¹: the observation that the size profile of the rarer of two closely related and similar species is in every case more irregular than that of the commoner species². This observation cannot be explained by assuming diffuse competition. Rather than contradicting our model, we suggest instead that the empirical evidence regarding niche breadth may be incomplete.

Our identification of focused competition could be related to our comparison of congeners, an idea that is supported by evidence that, when phylogenetic relatedness is taken into account, functional variation in several physiological traits is predicted by characteristics other than presumed 'functional' (niche) types^{3,4}. Woody communities are composed, on average, of roughly 30% congeners⁵, and we may well have identified a general process of forest dynamics.

Second, absolute dating of fluctuations is not crucial to our model, or for discarding others. Our conversion from size to age profiles defines equivalent time spans over which to compare fluctuations within individual pairs. No other model of coexistence predicts that the rarer species will be the superior competitor; the uniformity of that pattern in our data discounts alternative models.

As a counter-example to our method, Lusk derives fluctuations from separate profiles of age measured directly and age derived from diameter. The relationship between the two is to be expected in a population with little or no recruitment fluctuation, as is the case for the *Weinmannia racemosa* data⁶ used by Lusk. We have found that, where two-dimensional plots are available^{7,8}, mean diameter increases roughly linearly with age, and the variance of diameter for a given age also increases with age. With such a relationship and without recruitment fluctuations, the two measures of fluctuation would initially correspond, but over time this correspondence would degrade through statistical fluctuation alone, as in Lusk's figure. From his raw data, we calculate that the degree of fluctuation for these is about 0.5, a value that is comparable with that for the smoothest profile in our target species, the common *Bursera instabilis*¹, and is consistent with statistical

variation alone. The deviation values for our rarer species vary from about 3 to 5.

Baker and Wilson make the point that growth responses differ in different environmental conditions, thereby producing different relationships between age and diameter. But our method assumes that species with a large degree of ecological and evolutionary similarity will respond similarly to the same environmental conditions. We compared morphologically similar congeneric species, which in one instance were sister species, having first checked that the paired congeners had the same habitat use and shade tolerance^{2,9}. The compared species occurred together, with congener populations being interdigitated so that nearest neighbours could be either conspecifics or congeners. Baker and Wilson's "wind storms, fires or climatic anomalies" would be hard put to have different effects within our paired congeners.

By contrast, Baker and Wilson compare species in different genera, each of which is in a different tribe within a large family (Meliaceae; > 550 spp., > 50 genera¹⁰), signifying considerable genetic distance. The authors do not establish ecological comparability at the level that we did for our paired comparisons. From Baker and Wilson's Fig. 1, we estimate sample sizes of roughly 20 individuals for each species. For samples of this size, statistical variation in the measure of deviation could be as large as the values that they present.

We encourage identification of the limits of our assumptions regarding the diameter–age relationship; their definition will help to identify limits to documenting the Kelly–Bowler dynamic in nature. Unfortunately, a single example, even if it were relevant, cannot define such limits. Definition of the point where measurable similarity no longer usefully predicts unmeasurable similarity will require either better knowledge of the diameter–age relationship, or better analytical and simulation techniques. It would be a pity if Baker and Wilson's example were to dissuade ecologists from reasonable use of a historical record that is not otherwise readily available for many long-lived species.

Colleen Kelly*, Michael G. Bowler†

**Division of Biodiversity and Ecology, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, UK*

e-mail: c.k.kelly@soton.ac.uk

†Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK

1. Kelly, C. K. & Bowler, M. G. *Nature* **417**, 437–440 (2002).
2. Kelly, C. K. et al. *Ecol. Lett.* **4**, 618–627 (2001).
3. Kelly, C. K. J. *Vegetation Sci.* **7**, 417–424 (1996).
4. Kelly, C. K. & Woodward, F. I. J. *Ecol.* **83**, 509–515 (1995).
5. Gentry, A. H. *The Missouri Botanical Garden*. www.mobot.org/mobot/research/gentry/welcome.shtml (2003).
6. Lusk, C. H. & Ogden, J. *Ecol.* **80**, 379–393 (1992).
7. van Groenendael, J. M. et al. *J. Trop. Ecol.* **12**, 11–24 (1996).
8. Platt, W. J. et al. *Am. Nat.* **131**, 491–525 (1988).
9. Buckley, Y. M. & Kelly, C. K. *Plant Ecol.* (in the press).
10. Mabblerly, D. J. *The Plant Book* 2nd edn (Cambridge Univ. Press, Cambridge, 1997).

Wnt-mediated axon guidance via the *Drosophila* Derailed receptor

Shingo Yoshikawa*, Randall D. McKinnon*†, Michelle Kokel* & John B. Thomas*

* Molecular Neurobiology Laboratory, Salk Institute for Biological Studies, PO Box 85800, San Diego, California 92186, USA

† Departments of Surgery/Molecular Genetics Microbiology, Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, New Jersey 08854, USA

In nervous systems with bilateral symmetry, many neurons project axons across the midline to the opposite side. In each segment of the *Drosophila* embryonic nervous system, axons that display this projection pattern choose one of two distinct tracts: the anterior or posterior commissure. Commissure choice is controlled by Derailed, an atypical receptor tyrosine kinase expressed on axons projecting in the anterior commissure. Here we show that Derailed keeps these axons out of the posterior commissure by acting as a receptor for Wnt5, a member of the Wnt family of secreted signalling molecules. Our results reveal an unexpected role in axon guidance for a Wnt family member, and show that the Derailed receptor is an essential component of Wnt signalling in these guidance events.

The growth cones of developing neurons are guided to their targets by attractive and repulsive cues in the extracellular environment. Specific receptors on the growth cones recognize these cues and transduce signals that ultimately lead to changes in direction of growth. The best understood of these cues and their axonal receptors are involved in guidance of the large number of axons that project across the midline to the opposite side of the central nervous system (CNS)¹. Distinct groups of cells at the midline divide the two halves of the CNS and have a critical role in axon guidance. These cells, termed midline glia in *Drosophila*, secrete diffusible factors, the Netrins, capable of attracting contralaterally projecting axons^{2–4}. They also secrete a repellent factor Slit, which together with its receptor Roundabout (Robo) and an intracellular sorting factor that modulates the delivery of Robo to the cell surface, controls whether or not axons will cross the midline^{5–11}.

Once axons commit to crossing the midline, they do not do so randomly. Instead, they follow particular tracts. In each segment of the *Drosophila* embryonic ventral nerve cord, crossing axons choose one of two commissural tracts, either the anterior or posterior commissure (AC or PC, respectively), which connect the two sides. This choice of commissure is controlled in part by the Derailed (Drl) guidance receptor. Drl is expressed on the growth cones and axons of all neurons that project through the AC, and seems to act as a receptor for a repellent factor in the PC¹². In *drl* mutants, AC axons abnormally cross in the PC of many segments. Conversely, misexpression of Drl in PC neurons switches their axonal projections to the AC. Thus, Drl is both necessary and sufficient for axons to cross the midline in the AC. The behavioural phenotypes of *drl* mutants suggest that at least some of the neurons that require Drl for their guidance fail to make synaptic connections essential for coordinated locomotion and learning and memory^{13,14}.

Drl is a member of the RYK subfamily of atypical receptor tyrosine kinases (RTKs). All members of this subfamily have unusual, but highly conserved amino acid substitutions in their kinase domains plus relatively short extracellular domains devoid of motifs commonly found in other RTKs^{13,15–17}. Consistent with the unusual amino acid substitutions, the kinase domain of RYK family members appears to lack catalytic activity^{18–21}. However, whereas catalytic activity of Drl has been shown to be dispensable, its cytoplasmic domain is required to dictate commissure choice, suggesting that Drl transduces a signal within growth cones in an unconventional manner, perhaps together with another catalytically active kinase²⁰. The extracellular domain of each RYK family member contains a Wnt inhibitory factor (WIF) domain²². The

WIF domain of other molecules has been shown to bind to and inhibit the function of members of the Wnt family of secreted signalling molecules²³, raising the possibility that members of the RYK receptor family, including Drl, bind to Wnt proteins. The Wnt family is large, consisting of seven members in *Drosophila* and 19 in humans, and is involved in a diverse array of developmental events. Wnt proteins have well-established roles in early cell fate decisions and embryonic patterning²⁴, but have also been implicated in synaptic remodelling and terminal arborization within the developing CNS^{25–27}, as well as in regulating planar cell polarity by virtue of the phenotypes of mutations in Frizzled (Fz), one of the *Drosophila* members of the Fz family of Wnt receptors^{28,29}.

Genetic interaction between Drl and Wnt5

To identify components of the Drl signalling pathway, we carried out a genetic screen for mutations that suppress the ability of Drl to switch axons to the AC when misexpressed by PC neurons. We first screened a set of chromosomal deletions covering approximately 80% of the *Drosophila* genome. One of the deletions that showed strong dominant suppression of the PC-to-AC switching activity of Drl is *Df(1)N19*, a deletion that removes the X chromosome interval 17A1 to 18A2. By testing a series of overlapping deletions within the *Df(1)N19* region, we narrowed the interval to 17B. One of the genes in this interval is *wnt5* (also called *Dwnt3*), a member of the Wnt gene family in *Drosophila*. *wnt5* is a single-exon gene encoding an

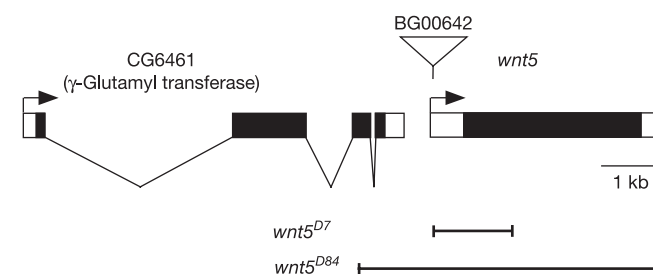


Figure 1 Schematic diagram of the *wnt5* genomic region. Wnt5 is encoded by a single-exon gene. The P element BG00642 is inserted in the 5' UTR of *wnt5*, 75 bp downstream of the transcriptional start site. Below are the extents of deletions generated by imprecise excision of BG00642. *wnt5*^{D7} is a 1,251-bp deletion that removes the first 261 amino acids of the Wnt5 gene product (coding region shaded in black). *wnt5*^{D84} is a 5,249-bp deletion that removes the entire *wnt5* coding region plus part of the 3' end of the adjacent gene, CG6461, encoding a member of a family of γ -glutamyl transferases.

unusually large Wnt protein of 1,004 amino acids with a unique amino-terminal domain that seems to be proteolytically cleaved³⁰, followed by the Wnt domain common to all members of the family^{31,32}.

Given the possibility that Drl might interact with Wnt proteins by means of its WIF domain, we generated mutations specifically in the *wnt5* gene to test whether reduction of Wnt5 itself is responsible for the suppression observed with the larger chromosomal deletions. We identified a P element transposon, BG00642, from the Berkeley Drosophila Genome Project inserted in the 5'-untranslated region (5' UTR) of *wnt5*, and mobilized it to generate deletions of the *wnt5* coding region (Fig. 1). We recovered a deletion, *wnt5*^{D7}, that removes the first 261 amino acids of the Wnt5 protein but does not affect either adjacent gene. In addition, we recovered a larger deletion, *wnt5*^{D84}, that removes the entire *wnt5* coding region plus part of the 3' end of the adjacent gene encoding a member of a family of γ -glutamyl transferases. Both *wnt5*^{D7} and *wnt5*^{D84} abolish Wnt5 expression (see below), and the phenotypes of *wnt5*^{D7} and *wnt5*^{D84} homozygotes, as well as *wnt5*^{D7}/*wnt5*^{D84} individuals, are indistinguishable. Thus, *wnt5*^{D7} acts as a null allele. *wnt5*^{D7} and *wnt5*^{D84} homozygotes are viable and fertile, but similar to *drl* mutants, adults are uncoordinated.

We first tested whether mutations in *wnt5* could suppress the ability of Drl to switch axons to the AC. Using the Gal4/UAS transactivation system³³, we misexpressed Drl in PC neurons with *eagle-GAL4* (*eg-GAL4*)³⁴. This driver expresses Gal4 in a sufficiently small subset of neurons so that we could follow unambiguously their axonal projections with a *UAS-tau-myc-green fluorescent protein* (*GFP*) axon-targeted reporter transgene¹². *eg-GAL4* drives expression in two small clusters of Eg interneurons in each hemisegment, both of which project axons across the midline (Fig. 2a). One of the clusters projects in the PC and the other in the AC. At the midline, the axons from homologous clusters on either side of each segment fasciculate with one another, forming two distinct axon bundles, one within each of the commissures. When forced to misexpress Drl using a *UAS-drl* transgene, Eg PC neurons switched their projections to the AC in all segments (Fig. 2b and Table 1), whereas Eg AC neurons were unaffected. In 92% of segments, every Eg PC axon was switched to the AC, whereas in 8% of segments some axons remained within the PC. Misexpression of Drl using the same transgenes, but in a *wnt5*/+ heterozygous background, resulted in significantly fewer PC-to-AC switched axons: 34% of segments had all axons switched, 34% had some switched and 32% had none switched (Table 1). Notably, when Drl was misexpressed

in a *wnt5* hemizygous or homozygous mutant background, its ability to switch Eg PC neurons to the AC was completely abolished (Fig. 2c and Table 1). Thus, Drl requires Wnt5 to switch axons to a different commissure, suggesting that Wnt5 is an essential component of the Drl signalling pathway.

To examine the specificity of the genetic interaction between *wnt5* and *drl*, we tested two additional members of the *Drosophila* Wnt family, both of which are expressed in the embryonic CNS. In contrast to *wnt5*, mutations in either *wingless* (*wg*) or *wnt4* failed to show any suppression of Drl-mediated axon switching (Table 1), indicating that among these three Wnt family members, genetic interaction with *drl* is specific to *wnt5*.

Wnt5 expression in the CNS

In situ hybridization of *wnt5* probes to wild-type embryos revealed that *wnt5* messenger RNA expression in the CNS commences at stage 12, a point in development when differentiating neurons begin to extend axons, and continues throughout embryogenesis. We detected high levels of *wnt5* in subsets of neurons restricted to the posterior half of each segment and low levels in neurons located more anteriorly in the segment (Fig. 3a). The neurons expressing high levels of *wnt5* lie adjacent to and ventral to the PC in each segment (Fig. 3b, c). Although the precise identity of these neurons is unknown, their proximity to the PC and the fact that most of the CNS neurons project axons across the midline suggest that many, perhaps all, of the *wnt5*-expressing neurons project axons through the PC.

To examine the extracellular distribution of Wnt5 protein, we stained live embryos with an antibody raised against a unique region of the protein N-terminal to the Wnt domain³⁰. We detected staining on the major axonal tracts within the CNS, with the highest levels on the two commissures (Fig. 3d), a pattern similar to that described previously³⁰. Staining is abolished in *wnt5* mutants (Fig. 3e), demonstrating the specificity of the antibody. Given that *wnt5* mRNA is expressed predominantly by subsets of neurons located posteriorly within the segment, Wnt5 apparently either diffuses to the AC or is picked up by AC growth cones and axons as they project towards the midline. Whether the proteolytic processing of Wnt5 observed in cultured cells is involved in its distribution *in vivo* is not known, as the anti-Wnt5 antibody recognizes both the unprocessed and processed forms of the protein (relative molecular mass (*M_r*) of 140K and 80K, respectively)³⁰. Similarly, it is unknown whether all of the Wnt5 recognized by the antibody is biologically active, as processing may be required for activity.

Wnt5 function in midline crossing

If Wnt5 were a component of the Drl signalling pathway, then loss-of-function mutations in *wnt5* might be expected to exhibit *drl*-like mutant phenotypes. Using an antibody that labels all CNS axons, we found that *wnt5* mutant embryos, similar to *drl* mutants, have disorganized commissures (Fig. 4b, c). In many segments commis-

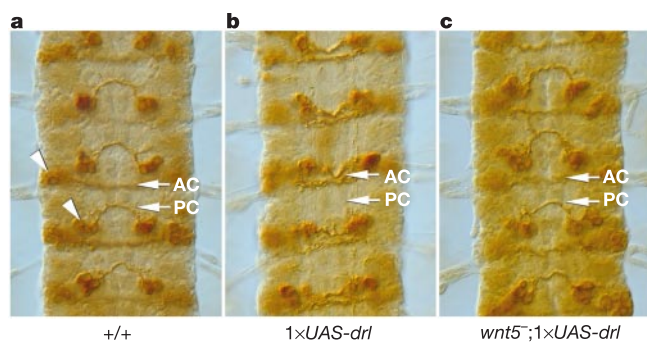


Figure 2 *wnt5* suppression of Drl-mediated axon switching. **a**, In wild-type (+/+) embryos, clusters of Eg neurons (arrowheads) in each hemisegment project axons across the midline, forming two bundles, one in the AC and one in the PC. **b**, When forced to misexpress Drl from a *UAS-drl* transgene, nearly all Eg PC axons cross the midline in the AC, forming a distinct ectopic bundle. **c**, In a *wnt5* mutant background, Drl fails to switch any of the Eg PC axons to the AC. See Table 1 for quantification. Genotypes: **a**, *UAS-tau-myc-GFP*/+; *eg-GAL4*/+; **b**, *UAS-drl*, *UAS-tau-myc-GFP*/+; *eg-GAL4*/+; **c**, *wnt5*^{D7}/*Y*; *UAS-drl*, *UAS-tau-myc-GFP*/+; *eg-GAL4*/+.

Table 1 Suppression of Drl-mediated PC-to-AC axon switching

Mutant genotype	Percentage of segments			N
	All axons switched	Some axons switched	No axons switched	
No <i>UAS-drl</i>	0	0	100	(N = 100)
1 × <i>UAS-drl</i>	92	8	0	(N = 100)
<i>wnt5</i> /+; 1 × <i>UAS-drl</i>	34	34	32	(N = 96)
<i>wnt5</i> / <i>wnt5</i> ; 1 × <i>UAS-drl</i>	0	0	100	(N = 98)
<i>wg</i> /+; 1 × <i>UAS-drl</i>	88	12	0	(N = 100)
<i>wnt4</i> /+; 1 × <i>UAS-drl</i>	93	7	0	(N = 100)
<i>fz fz2</i> /+; 1 × <i>UAS-drl</i>	95	5	0	(N = 79)
<i>dsh</i> / <i>Y</i> ; 1 × <i>UAS-drl</i>	94	6	0	(N = 100)
3 × <i>UAS-drl</i> Δ i	0	0	100	(N = 73)

All embryos carried the *eg-GAL4* driver plus a *UAS-tau-myc-GFP* transgene to visualize axonal projections. The same *UAS-drl* transgene was used.

tures appear irregular, and there are often abnormal axonal projections between the AC and PC. To determine whether these abnormal projections arise from AC axons projecting to the PC or vice versa, we labelled subsets of AC and PC axons in *wnt5* mutants with marker lines previously used for analysing *drl* mutants³⁵. In all cases, we found that AC axons either wander from the AC into the PC or project entirely through the PC, whereas projections of PC axons appear unaltered. These defects are similar to those seen in *drl* mutants¹² (Fig. 4e). For example, in *wnt5* mutants assayed with a *P{tau-lacZ}* marker for AC axons, 80% of segments displayed abnormal projections of AC axons into the PC (Fig. 4f). We obtained similar results using *Sema2b-tau-myc*, another marker for a subset of AC neurons³⁶. In contrast to AC axons, in no segments did PC axons, as assayed with a *P{tau-lacZ}* marker for the PC, project abnormally into the AC (Fig. 4g). Thus, Wnt5, similar to Drl, is required for proper projection of AC axons across the midline of the CNS.

We previously proposed that Drl functions to keep axons in the AC by acting as a guidance receptor for a repellent ligand in the PC¹². The high levels of expression by neurons associated with the PC is consistent with Wnt5 acting as such a repulsive ligand. To examine whether Wnt5 is capable of repelling Drl-expressing axons, we misexpressed it at the midline and assayed the effect on crossing axons. Misexpression of Wnt5 in midline glia using the *sim-GAL4*

driver³⁷ caused a marked reduction or complete loss of AC in 43% of segments ($N = 150$), but had no discernible effect on the PC (Fig. 4h). The affected AC axons appeared to either stall or project ipsilaterally within the longitudinal connectives. We interpret this phenotype as repulsion of the Drl-expressing AC axons from the ectopic source of Wnt5 at the midline. To determine whether this loss of the AC was dependent on Drl, we misexpressed Wnt5 using the identical combination of transgenes, but in a *drl* homozygous mutant background. Elimination of Drl completely suppressed the loss of the AC (Fig. 4i), restoring it in every segment, although as expected, we detected abnormal axonal projections between the AC and PC due to the loss of Drl. Thus, in the absence of Drl, axons are insensitive to Wnt5, a feature that may explain the tight spatial regulation of the Drl receptor during development. Drl is normally expressed on growth cones and axons as they project through the AC, but is rapidly downregulated once these growth cones leave the commissure and begin to project in the longitudinal connectives on the contralateral side¹². This downregulation of Drl may be required to allow further extension of the AC growth cones as they traverse regions of repellent Wnt5 in the connectives, similar to the downregulation of Robo allowing axons to traverse the Slit-expressing midline^{6,11}.

Physical interaction between Drl and Wnt5

The above results suggest that Drl is a receptor for Wnt5. To test for binding of Drl to Wnt5, we first examined the *in vivo* binding of Drl-Fc, a soluble probe consisting of the Drl extracellular domain epitope-tagged with the human immunoglobulin- γ (IgG) Fc fragment. In wild-type embryos, we detected Drl-Fc binding to the PC and to regions at the intersection of the PC and the longitudinal connectives (Fig. 5a). In *wnt5* mutant embryos, binding of Drl-Fc was abolished (Fig. 5b). Conversely, when we misexpressed Wnt5 by heat-shocking late-stage embryos carrying a *heat shock-wnt5* (*hs-wnt5*) transgene, followed by incubation with Drl-Fc, binding was markedly expanded to include all axon tracts in the CNS (Fig. 5c). The observation that Drl-Fc labelled only the PC in wild-type embryos, whereas Wnt5 is present on both commissures (compare Figs 3d and 5a), suggests that Wnt5 in the AC is bound to endogenous Drl present on the AC growth cones and axons, and that this interaction may block Drl-Fc access to Wnt5. Consistent with this, in *drl* mutants Drl-Fc labelled both the AC as well as the PC (Fig. 5e).

To test for specificity of the Drl-Fc–Wnt5 interaction, we examined Drl-Fc binding in embryos misexpressing high levels of either Wg or Wnt4. We detected no expansion of Drl-Fc binding in these embryos (Fig. 5d shows *hs-wg*). These results are consistent with the lack of genetic interaction between *drl* and either *wg* or *wnt4*, and suggest further that Drl receptor binding is restricted to certain members, perhaps a single member, of the Wnt family.

We next tested whether Drl-Fc could bind Wnt5 in solution. Whole embryo extracts were prepared from wild-type and *wnt5* homozygous mutant embryos, then incubated with Drl-Fc coupled to agarose beads. As a control, extracts were also incubated with beads coupled with human IgG heavy chain alone. Bound material was analysed by SDS–polyacrylamide gel electrophoresis (PAGE) and immunoblotted with the anti-Wnt5 antibody. We found that Drl-Fc was able to co-precipitate both the unprocessed and the proteolytically processed forms of Wnt5, as evidenced by the presence of 140K and 80K bands from wild-type extracts, but not from *wnt5* mutant extracts (Fig. 5f). Consistent with the lack of Drl-Fc binding to misexpressed Wg *in vivo*, Drl-Fc did not co-precipitate Wg, as assayed by immunoblotting with an anti-Wg antibody (not shown). Although both forms of Wnt5 present in the extracts are capable of binding to Drl-Fc, it is not known, *in vivo*, whether both actually have access to the Drl receptor. For example, the 140K form may not be efficiently secreted, as has been observed in cultured cells³⁰. However, regardless of the *in vivo* distribution of

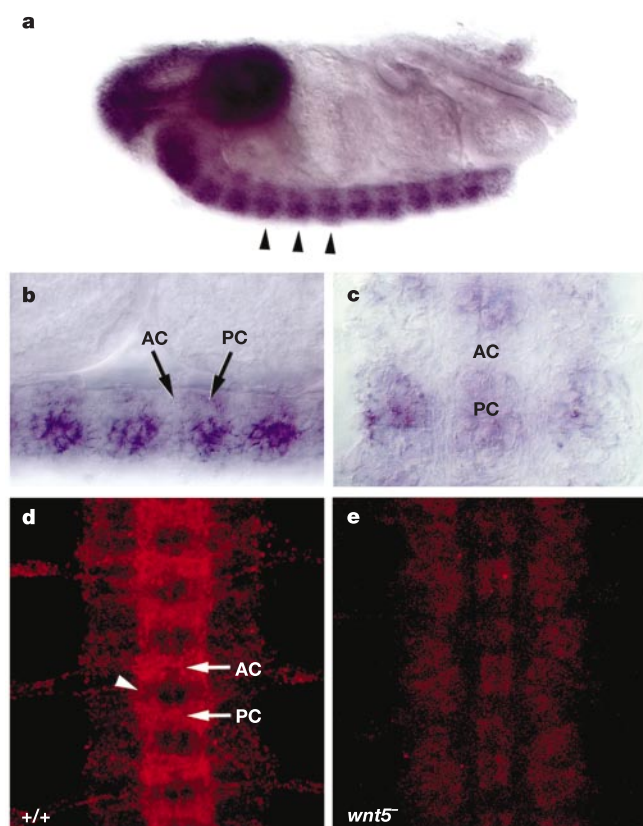


Figure 3 Wnt5 CNS expression. **a–c**, *In situ* hybridization of *wnt5* antisense RNA probes. Neurons within the posterior half of each segment of the ventral nerve cord express high levels of *wnt5* (arrowheads in **a** indicate three adjacent segments of a stage 16 embryo viewed sagittally). Higher magnification (**b**) shows *wnt5*-expressing cells lying ventral to the PC, which is visualized in cross-section. Ventral view of a stage 14 embryo (**c**) shows *wnt5*-expressing neurons adjacent to the PC. **d, e**, Immunofluorescence staining of stage 15 embryos with anti-Wnt5 antibodies. In wild type (**d**), high levels of Wnt5 are present on the commissures, with lower levels on the connectives (arrowhead). Staining is abolished in *wnt5*^{D7} mutant embryos (**e**).

the two forms of Wnt5, this result, together with the genetic evidence, indicates that Wnt5 is the ligand for Drl.

Guidance mechanisms controlling midline crossing

The fact that Wnt proteins are known to signal through Fz receptors raises the possibility that Drl might not be acting as a 'classical' guidance receptor, but as a co-receptor for Wnt5, modulating its signalling through one or more of the Fz proteins in a manner similar to that proposed for Arrow/LRP6 (refs 38, 39). For example, binding of Wnt5 by Drl might modify Wnt5 signalling through Fz and/or Fz2, the two *Drosophila* Fz family members expressed by embryonic CNS neurons^{40–42}. However, in contrast to *wnt5*, neither *fz fz2* double mutants, nor mutations in *dishevelled* (*dsh*)—a downstream Fz signalling component—show any effect on Drl-mediated axon switching (Table 1), and *drl/+; fz fz2/+* trans-heterozygous embryos do not show defects in midline crossing (not shown). Furthermore, interfering with Fz-mediated Wnt signalling by pan-neuronally expressing a dominant-negative form of Fz2, GPI-Dfz2 (ref. 43), causes no defects in midline crossing (not shown). Although these results do not rule out signalling through Fz proteins, they do advance the idea that Wnt5 might be signalling through the Drl receptor, a possibility consistent with the finding

that misexpression of Drl lacking its intracellular domain (*DrlΔi*) fails to switch any Eg axons, even when misexpressed at high levels from multiple transgenes (Table 1). In either event, whether Drl transduces the Wnt5 signal or modulates Wnt5 signalling through Fz proteins, it is an essential component of Wnt5 signalling in the guidance of axons across the midline.

In both *drl* and *wnt5* mutants many axons still project appropriately in the AC, suggesting that Wnt5 and Drl are part of a larger multi-component system to ensure proper sorting of axons as they cross the midline. For example, it seems probable that there are additional attractive cues for AC axons, and that once they are attracted to the AC, Drl functions to prevent them from entering the PC. In addition, there may be a similar mechanism for the guidance of PC axons, whose choice of commissure is unaffected in both *drl* and *wnt5* mutant embryos. Possibilities for additional molecules involved in commissure choice include other members of the Drl and Wnt families, some of which are expressed in the developing CNS (L. Santschi and J.B.T., unpublished).

Our results in *Drosophila* suggest that a similar receptor–ligand interaction between RYK and Wnt family members might be functioning in mammalian CNS development. Although nervous system phenotypes have not yet been described for the mouse

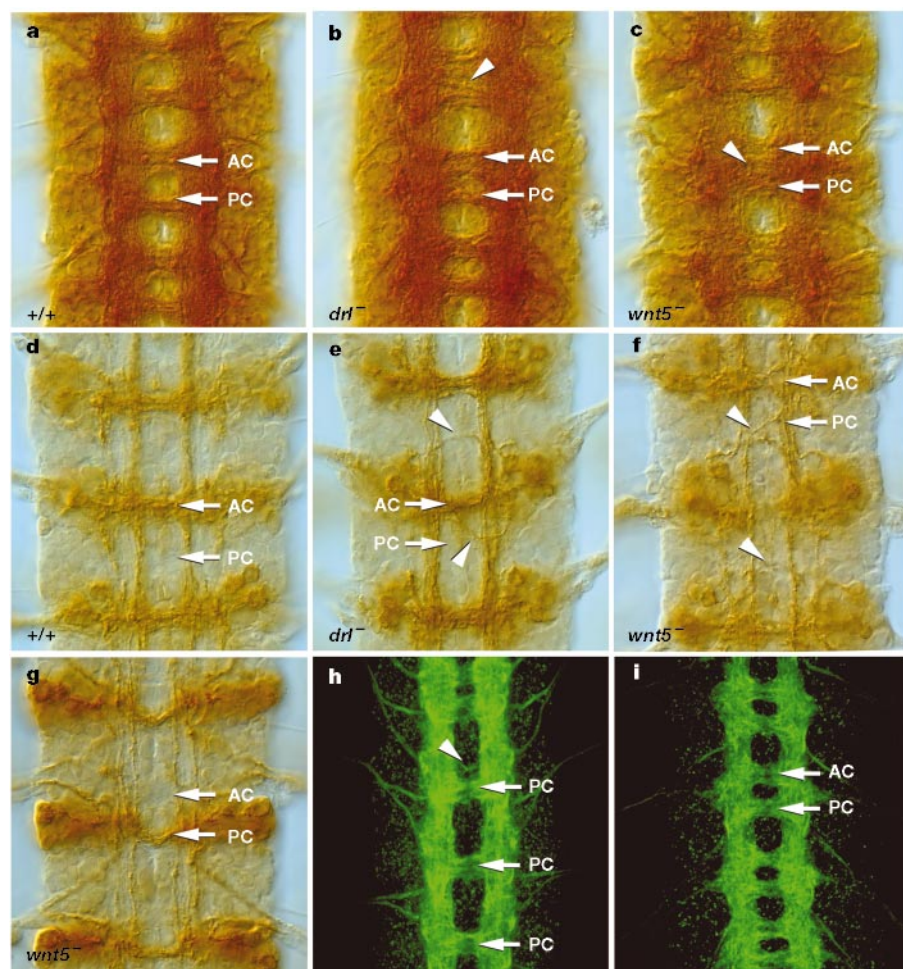


Figure 4 *wnt5* CNS phenotypes. **a–c**, Embryos immunostained for all axons. In contrast to wild-type embryos (**a**), *drl*^{R343} (**b**) and *wnt5*^{D7} (**c**) embryos have axons abnormally crossing between the AC and PC (arrowheads). **d–g**, Embryos labelled for subsets of AC axons (**d–f**) or PC axons (**g**). In contrast to wild-type (**d**) embryos, AC axons of *drl*^{R343} (**e**) and *wnt5*^{D7} (**f**) embryos project abnormally in the PC (arrowheads). PC axons in *wnt5*^{D7} (**g**) project normally. **h, i**, Embryos immunostained for all axons. In embryos

misexpressing Wnt5 by the midline glia (**h**), the AC is missing or is severely reduced (arrowhead), but in a *drl* mutant background (**i**) the AC is restored in all segments. Genotypes: **h**, *P[EPgy2]EY03178/Y; sim-GAL4₂/+; sim-GAL4₃/+*; **i**, *P[EPgy2]EY03178/Y; drl*^{R343}, *sim-GAL4₂/drl*^{R343}, *sim-GAL4₃/+*. (Subscripts indicate whether the *sim-GAL4* transgene occurs on the 2nd or 3rd chromosome.)

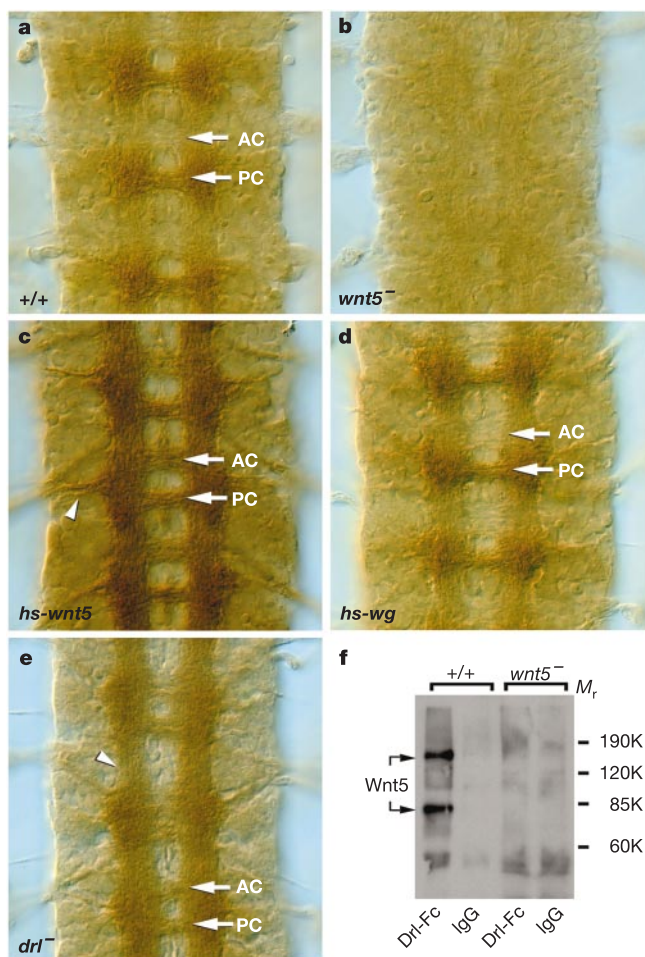


Figure 5 Drl binding to Wnt5. **a–e**, Embryos incubated with Drl-Fc. In wild-type (**a**) embryos, Drl-Fc binds to the PC and its junctions with the connectives. Binding is abolished in *wnt5*^{D7} mutants (**b**). *hs-wnt5* embryos misexpressing Wnt5 (**c**) show expansion of Drl-Fc binding to all axon tracts, including peripheral nerves (arrowhead), whereas *hs-wg* embryos misexpressing Wg (**d**) show no expansion. In *drl*^{R343} mutants (**e**), Drl-Fc binding is expanded to the AC and its junctions with the connectives (arrowhead). **f**, Drl-Fc immunoprecipitates Wnt5 from whole embryo lysates of wild-type but not *wnt5*^{D84} embryos. Control human IgG heavy chain fails to immunoprecipitate significant amounts of Wnt5.

knockouts of RYK and Wnt5a, the two mutants, although differing in severity, do display qualitatively similar skeletal defects^{21,44}, suggesting the possibility of an interaction. Within the mammalian CNS, Wnt proteins have been implicated in the guidance of commissural axons along the anterior–posterior axis of the spinal cord after they cross the midline (Y. Zou, personal communication). It will be of interest to test whether RYK has a role in these guidance events.

There are at least three distinct guidance mechanisms involved in midline crossing of contralaterally projecting axons within the *Drosophila* CNS. As in vertebrates, growing axons are attracted to the midline by diffusible cues such as Netrins acting through their receptor Frazzled/Dcc. Once there, the choice of whether or not to cross is controlled by Slit through its receptor Robo. Finally, as shown here, their choice of commissure is controlled by Wnt5 by means of its receptor Drl. □

Methods

Fly stocks and genetics

P-element insertion lines BG00642 and EY03178 were obtained from the Berkeley *Drosophila* Genome Project. BG00642 is a P{GT1} element inserted in the 5' UTR of *wnt5*,

75 base pairs (bp) downstream of the transcriptional start site. EY03178 is an EPgy2 element⁴⁵ inserted 23 bp upstream of the *wnt5* transcriptional start site and produces high levels of Gal4-dependent expression of Wnt5. Extents of the *wnt5*^{D7} and *wnt5*^{D84} deletions were confirmed by sequencing. In all experiments the genotypes of embryos were verified with marked balancer chromosomes *FM7,Act-GFP*, *CyO,wg-lacZ* and *TM3,ftz-lacZ*. Specific alleles used were: *drl*^{R343}, *drl*^{Red2} (refs 12, 13), *wg*^{en11}, *wg*^{J-17}, *wnt4*^{EMS23} (ref. 46), *fz*^{H51}, *fz*^{C1} (ref. 42), *dsh*³. Transgenes used were: *P{tau-lacZ}3.765* and *P{tau-lacZ}3.538* for labelling AC and PC axons, respectively³⁵; *Sema2b-tau-myc* (ref. 36), *UAS-tau-myc-GFP* (ref. 47), *UAS-drl* (ref. 12), *UAS-Dwnt4* (ref. 48), *hs-wnt5* (ref. 30), *hs-wg* (ref. 49), *UAS-GPI-Dfz2* (ref. 43), *eg-GAL4* (ref. 34), *sim-GAL4* (ref. 37), *elav-GAL4* (ref. 10).

Constructs

UAS-drlΔi was constructed by ligating a fragment of a *drl* complementary DNA¹³ encoding the extracellular and transmembrane domains plus the first 12 amino acids of the cytoplasmic domain to sequences encoding six copies of the c-Myc epitope, followed by cloning into pUAST. Expression and proper localization of DrlΔi to growth cones was confirmed by immunostaining for c-Myc. Drl-Fc was constructed by ligating a fragment of a *drl* cDNA¹³ encoding the extracellular domain to sequence encoding the human IgG Fc fragment, followed by cloning into the pFastBac vector (Gibco-BRL).

Histology

Immunohistochemistry and *in situ* hybridization were performed as described⁵⁰, with the exception that staining with the anti-Wnt5 antibody was carried out on live unfixed embryos, as fixation abolished binding of the antibody to Wnt5 protein. Anti-Myc antibody was used to stain for Tau-Myc-GFP. Staining with Drl-Fc was performed as previously described for Drl-Myc¹³. For Drl-Fc staining of *hs-wnt5*, *hs-wg* and *hs-GAL4/UAS-Dwnt4* embryos, we carried out a heat shock twice for 1 h each at 37 °C followed by incubation with Drl-Fc.

Solution binding assay

Baculovirus expression was performed as described for the Bac-to-Bac system (Gibco-BRL). For maximum protein expression, Hi5 cells were infected at about 70% confluence and the medium was collected after 6 days. To detect interaction between Drl-Fc and Wnt5, 100 μl of protein A agarose beads (Calbiochem) were incubated with 100 μg of mouse anti-human Fc antibody (Jackson Immuno Research Laboratory) in TBS (20 mM Tris-HCl, pH 7.5, 0.15 M NaCl) at room temperature for 2 h. After washing three times with 1 ml of TBS, 40 μl of the beads was incubated with 500 μl of conditioned medium containing either Drl-Fc or human IgG at room temperature for 2 h. The beads were washed three times with TBS and then incubated with extracts from either wild-type or *wnt5* mutant embryos with IP buffer (TBS with 1% Triton X-100) at 4 °C for 12 h, and washed ten times with 1 ml of IP buffer. Bound protein was released by boiling beads in 40 μl of SDS sample buffer and analysed by SDS-PAGE (8% gel) and western blotting with anti-Wnt5 or anti-Wg antibody.

Received 3 February; accepted 4 March 2003; doi:10.1038/nature01522.

Published online 16 March 2003.

- Dickson, B. J. Molecular mechanisms of axon guidance. *Science* **298**, 1959–1964 (2002).
- Kennedy, T. E., Serafini, T., de la Torre, J. R. & Tessier-Lavigne, M. Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell* **78**, 425–435 (1994).
- Harris, R., Sabatelli, L. M. & Seeger, M. A. Guidance cues at the *Drosophila* CNS midline: identification and characterization of two *Drosophila* Netrin/UNC-6 homologs. *Neuron* **17**, 217–228 (1996).
- Mitchell, K. J. *et al.* Genetic analysis of Netrin genes in *Drosophila*: Netrins guide CNS commissural axons and peripheral motor axons. *Neuron* **17**, 203–215 (1996).
- Seeger, M., Tear, G., Ferres, M. D. & Goodman, C. S. Mutations affecting growth cone guidance in *Drosophila*: genes necessary for guidance toward or away from the midline. *Neuron* **10**, 409–426 (1993).
- Kidd, T., Russell, C., Goodman, C. S. & Tear, G. Dosage-sensitive and complementary functions of roundabout and commissureless control axon crossing of the CNS midline. *Neuron* **20**, 25–33 (1998).
- Kidd, T. *et al.* Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. *Cell* **92**, 205–215 (1998).
- Kidd, T., Bland, K. S. & Goodman, C. S. Slit is the midline repellent for the robo receptor in *Drosophila*. *Cell* **96**, 785–794 (1999).
- Rajagopalan, S., Nicolas, E., Vivancos, V., Berger, J. & Dickson, B. J. Crossing the midline: roles and regulation of Robo receptors. *Neuron* **28**, 767–777 (2000).
- Simpson, J. H., Kidd, T., Bland, K. S. & Goodman, C. S. Short-range and long-range guidance by slit and its Robo receptors. Robo and Robo2 play distinct roles in midline guidance. *Neuron* **28**, 753–766 (2000).
- Keleman, K. *et al.* Comm sorts robo to control axon guidance at the *Drosophila* midline. *Cell* **110**, 415–427 (2002).
- Bonkowski, J. L., Yoshikawa, S., O'Keefe, D. D., Scully, A. L. & Thomas, J. B. Axon routing across the midline controlled by the *Drosophila* Derailed receptor. *Nature* **402**, 540–544 (1999).
- Callahan, C. A., Muralidhar, M. G., Lundgren, S. E., Scully, A. L. & Thomas, J. B. Control of neuronal pathway selection by a *Drosophila* receptor protein-tyrosine kinase family member. *Nature* **376**, 171–174 (1995).
- Moreau-Fauvarque, C., Taillebourg, E., Preat, T. & Dura, J. M. Mutation of linotte causes behavioral defects independently of pigeon in *Drosophila*. *Neuroreport* **13**, 2309–2312 (2002).
- Hovens, C. M. *et al.* RYK, a receptor tyrosine kinase-related molecule with unusual kinase domain motifs. *Proc. Natl Acad. Sci. USA* **89**, 11818–11822 (1992).
- Oates, A. C. *et al.* Embryonic expression and activity of doughnut, a second RYK homolog in *Drosophila*. *Mech. Dev.* **78**, 165–169 (1998).
- Halford, M. M., Oates, A. C., Hibbs, M. L., Wilks, A. F. & Stacker, S. A. Genomic structure and expression of the mouse growth factor receptor related to tyrosine kinases (Ryk). *J. Biol. Chem.* **274**, 7379–7390 (1999).

18. Katso, R. M., Russell, R. B. & Ganesan, T. S. Functional analysis of H-Ryk, an atypical member of the receptor tyrosine kinase family. *Mol. Cell Biol.* **19**, 6427–6440 (1999).
19. Savant-Bhonsale, S., Friese, M., McCoon, P. & Montell, D. J. A *Drosophila* derailed homolog, Doughnut, expressed in invaginating cells during embryogenesis. *Gene* **231**, 155–161 (1999).
20. Yoshikawa, S., Bonkowski, J. L., Kokel, M., Shyn, S. & Thomas, J. B. The derailed guidance receptor does not require kinase activity *in vivo*. *J. Neurosci.* **21**, RC119 (2001).
21. Halford, M. M. *et al.* Ryk-deficient mice exhibit craniofacial defects associated with perturbed Eph receptor crosstalk. *Nature Genet.* **25**, 414–418 (2000).
22. Pathy, L. The WIF module. *Trends Biochem. Sci.* **25**, 12–13 (2000).
23. Hsieh, J. C. *et al.* A new secreted protein that binds to Wnt proteins and inhibits their activities. *Nature* **398**, 431–436 (1999).
24. Wodarz, A. & Nusse, R. Mechanisms of Wnt signaling in development. *Annu. Rev. Cell Dev. Biol.* **14**, 59–88 (1998).
25. Hall, A. C., Lucas, F. R. & Salinas, P. C. Axonal remodeling and synaptic differentiation in the cerebellum is regulated by WNT-7a signaling. *Cell* **100**, 525–535 (2000).
26. Krylova, O. *et al.* WNT-3, expressed by motoneurons, regulates terminal arborization of neurotrophin-3-responsive spinal sensory neurons. *Neuron* **35**, 1043–1056 (2002).
27. Packard, M. *et al.* The *Drosophila* Wnt, wingless, provides an essential signal for pre- and postsynaptic differentiation. *Cell* **111**, 319–330 (2002).
28. Bhanot, P. *et al.* A new member of the frizzled family from *Drosophila* functions as a Wingless receptor. *Nature* **382**, 225–230 (1996).
29. Vinson, C. R. & Adler, P. N. Directional non-cell autonomy and the transmission of polarity information by the frizzled gene of *Drosophila*. *Nature* **329**, 549–551 (1987).
30. Fradkin, L. G., Noordermeer, J. N. & Nusse, R. The *Drosophila* Wnt protein DWnt-3 is a secreted glycoprotein localized on the axon tracts of the embryonic CNS. *Dev. Biol.* **168**, 202–213 (1995).
31. Eisenberg, L. M., Ingham, P. W. & Brown, A. M. Cloning and characterization of a novel *Drosophila* Wnt gene, Dwnt-5, a putative downstream target of the homeobox gene distal-less. *Dev. Biol.* **154**, 73–83 (1992).
32. Russell, J., Gennissen, A. & Nusse, R. Isolation and expression of two novel Wnt/wingless gene homologues in *Drosophila*. *Development* **115**, 475–485 (1992).
33. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
34. Dittrich, R., Bossing, T., Gould, A. P., Technau, G. M. & Urban, J. The differentiation of the serotonergic neurons in the *Drosophila* ventral nerve cord depends on the combined function of the zinc finger proteins Eagle and Hucklebein. *Development* **124**, 2515–2525 (1997).
35. Callahan, C. A. & Thomas, J. B. Tau- β -galactosidase, an axon-targeted fusion protein. *Proc. Natl Acad. Sci. USA* **91**, 5972–5976 (1994).
36. Rajagopalan, S., Vivancos, V., Nicolas, E. & Dickson, B. J. Selecting a longitudinal pathway: Robo receptors specify the lateral position of axons in the *Drosophila* CNS. *Cell* **103**, 1033–1045 (2000).
37. Golembo, M., Raz, E. & Shilo, B. Z. The *Drosophila* embryonic midline is the site of Spitz processing, and induces activation of the EGF receptor in the ventral ectoderm. *Development* **122**, 3363–3370 (1996).
38. Wehrli, M. *et al.* arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407**, 527–530 (2000).
39. Tamai, K. *et al.* LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530–535 (2000).
40. Bhat, K. M. frizzled and frizzled 2 play a partially redundant role in wingless signaling and have similar requirements to wingless in neurogenesis. *Cell* **95**, 1027–1036 (1998).
41. Bhanot, P. *et al.* Frizzled and Dfrizzled-2 function as redundant receptors for Wingless during *Drosophila* embryonic development. *Development* **126**, 4175–4186 (1999).
42. Chen, C. M. & Struhl, G. Wingless transduction by the Frizzled and Frizzled2 proteins of *Drosophila*. *Development* **126**, 5441–5452 (1999).
43. Cadigan, K. M., Fish, M. P., Rulifson, E. J. & Nusse, R. Wingless repression of *Drosophila* frizzled 2 expression shapes the Wingless morphogen gradient in the wing. *Cell* **93**, 767–777 (1998).
44. Yamaguchi, T. P., Bradley, A., McMahon, A. P. & Jones, S. A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* **126**, 1211–1223 (1999).
45. Mata, J., Curado, S., Ephrussi, A. & Rorth, P. Tribbles coordinates mitosis and morphogenesis in *Drosophila* by regulating string/CDC25 proteolysis. *Cell* **101**, 511–522 (2000).
46. Cohen, E. D. *et al.* DWnt4 regulates cell movement and focal adhesion kinase during *Drosophila* ovarian morphogenesis. *Dev. Cell* **2**, 437–448 (2002).
47. Callahan, C. A., Yoshikawa, S. & Thomas, J. B. Tracing axons. *Curr. Biol.* **8**, 582–586 (1998).
48. Gieseler, K. *et al.* Antagonist activity of DWnt-4 and wingless in the *Drosophila* embryonic ventral ectoderm and in heterologous *Xenopus* assays. *Mech. Dev.* **85**, 123–131 (1999).
49. Noordermeer, J., Johnston, P., Rijsewijk, F., Nusse, R. & Lawrence, P. A. The consequences of ubiquitous expression of the wingless gene in the *Drosophila* embryo. *Development* **116**, 711–719 (1992).
50. Thor, S. & Thomas, J. B. The *Drosophila* *islet* gene governs axon pathfinding and neurotransmitter identity. *Neuron* **18**, 397–409 (1997).

Acknowledgements We thank G. Struhl, R. Nusse, E. Wilder, B. Dickson and the Bloomington Stock Center for stocks; L. Fradkin and J. Noordermeer for the Wnt5 antibody, stocks and for sharing results prior to publication; and G. Lemke, L. Santschi and M. Boyle for discussion and critical reading of the manuscript. This work was supported by an NRSA fellowship to M.K. and grants from the NIH to R.D.M. and J.B.T. R.D.M. is a member of the Cancer Institute of New Jersey.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.B.T. (e-mail: jthomas@salk.edu).

Cold ions in the hot plasma sheet of Earth's magnetotail

Kanako Seki*, Masafumi Hirahara†, Masahiro Hoshino‡, Toshio Terasawa‡, Richard C. Elphic§, Yoshifumi Saito||, Toshifumi Mukai||, Hajime Hayakawa||, Hirotsugu Kojima¶ & Hiroshi Matsumoto¶

* Solar-Terrestrial Environment Laboratory, Nagoya University, Honohara 3-13, Toyokawa, Aichi 442-8507, Japan

† Department of Physics, Rikkyo University, Toshima, Tokyo 171-8501, Japan

‡ Department of Earth and Planetary Science, University of Tokyo, Tokyo 113-0033, Japan

§ NIS-1, MS D466, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

|| Institute of Space and Astronautical Science, Sagami-hara, Kanagawa 229-8510, Japan

¶ Radio Science Center for Space and Atmosphere, Kyoto University, Kyoto 611-0011, Japan

Most visible matter in the Universe exists as plasma. How this plasma is heated, and especially how the initial non-equilibrium plasma distributions relax to thermal equilibrium (as predicted by Maxwell–Boltzman statistics), is a fundamental question in studies of astrophysical^{1–3} and laboratory plasmas^{4,5}. Astrophysical plasmas are often so tenuous that binary collisions^{2,3} can be ignored, and it is not clear how thermal equilibrium develops for these ‘collisionless’ plasmas. One example of a collisionless plasma is the Earth’s plasma sheet⁶, where thermalized hot plasma with ion temperatures of about 5×10^7 K has been observed⁷. Here we report direct observations of a plasma distribution function during a solar eclipse, revealing cold ions in the Earth’s plasma sheet in coexistence with thermalized hot

ions. This cold component cannot be detected by plasma sensors on satellites that are positively charged in sunlight, but our observations in the Earth’s shadow show that the density of the cold ions is comparable to that of hot ions. This high density is difficult to explain within existing theories^{8–10}, as it requires a mechanism that permits half of the source plasma to remain cold upon entry into the hot turbulent plasma sheet.

Plasma is composed of an assembly of charged particles and is often referred to as the fourth state of matter because of its distinctive collective behaviour while maintaining charge neutrality. Plasma is quite common in the universe: most celestial objects and the interstellar medium are in the plasma state. A special feature of the plasma state is its wide variety of wave modes that enable effective energy transfer between ions, electrons and wave fields through various microscopic and macroscopic instabilities. This energization via waves can contribute to thermalization of the plasma, even if binary collision frequencies between constituent particles are much lower than the timescale of phenomena of interest.

In the nightside region of near-Earth space, a large-scale region called the plasma sheet (Fig. 1a) is formed through interaction between the intrinsic terrestrial magnetic field and the solar wind (a supersonic plasma flowing from the Sun). The plasma sheet lies between the northern and southern lobe regions characterized by strong magnetic field and cold (a few tens of electronvolts, eV) plasma. The cold plasma from the lobes becomes energized at the boundary layer between the plasma sheet and lobes, thereby forming the hot plasma sheet. The existence of this hot plasma sheet is thus observational evidence for the creation of thermalized hot plasma without collisions. Stimulated by the observations, plasma physicists have proposed a variety of acceleration and heating mechanisms for ions and electrons in the boundary layers of the plasma sheet^{8–10}. While the relative importance of each heating mechanism remains to be established, these theories predict that the originally cold source plasma cannot remain cold in the plasma

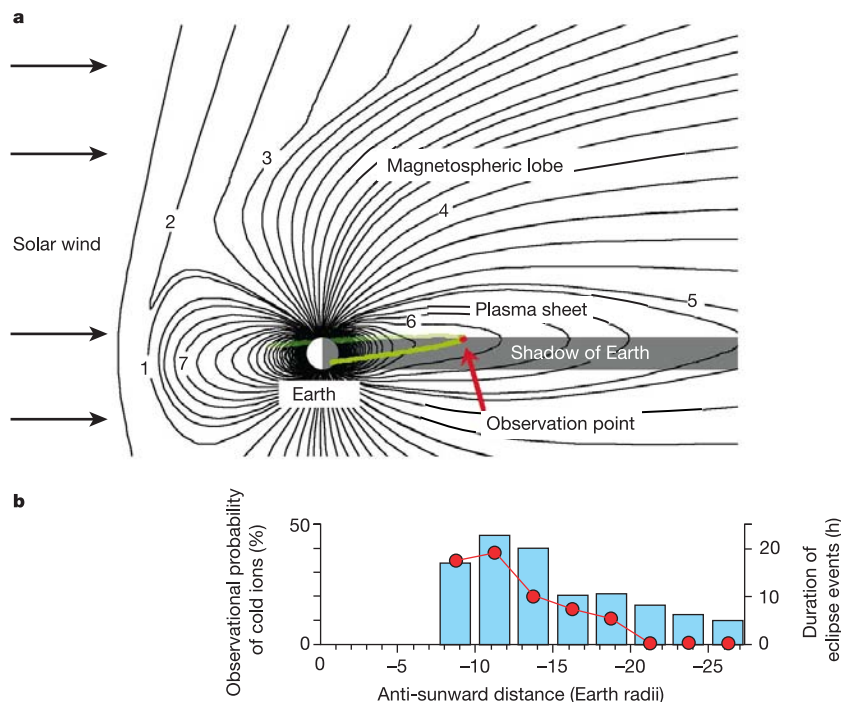


Figure 1 Location of the 19 July 1997 event and the statistical occurrence probability. **a**, Projection of GEOTAIL orbit onto the meridional plane (green line), together with typical magnetic field configuration during periods of southward interplanetary magnetic field. The Earth is indicated with a circle, the left-hand side of which corresponds to the direction towards the Sun. The red point shows the position where GEOTAIL entered the

shadow of the Earth, that is, the solar umbra. The numbered field lines illustrate the succession of configurations of a geomagnetic field line during magnetospheric convection. **b**, Probability of cold-ion detection in the plasma sheet from GEOTAIL eclipse data in 1995–1998 (red circles) together with total duration of eclipse events per bin of distance (blue bars) is shown as a function of distance from the Earth.

sheet. This prediction is widely accepted, because it has been supported by observations such as extensive wave activity¹¹, slow-mode shock structure¹², and ion heating¹³ in the plasma-sheet boundary layers.

This agreement between theoretical expectation and past observations, however, has a serious pitfall: a population of cold ions coexisting with the hot population has not been considered or observed. This is because the cold population is generally 'invisible' to plasma sensors aboard satellites in sunlight. A spacecraft in sunlight usually undergoes positive charging owing to photoelectron emission from its sunlit surfaces: this charging prevents positive ions with energies below the spacecraft potential from reaching detectors, and hence lowers the detection probability, especially at low energies. The difficulty of measuring cold ions disappears if a spacecraft enters the shadow of the Earth, because the vehicle undergoes negative charging in darkness and will detect ambient cold positive ions attracted by the spacecraft's negative charge. Using ion observations during passage through the solar umbra, we show here observational evidence of the existence of a significant cold-ion population in the hot plasma sheet.

The GEOTAIL spacecraft entered the northern edge of the shadow of the Earth at a position of about 9 Earth radii from the Earth on 19 July 1997 ('Observation point' in Fig. 1a). The geomagnetic conditions during the event were quiet throughout, following a series of geomagnetic substorms which ended at about 13:30 UT. The solar umbra (penumbra) began from 14:09 (14:03) UT, and lasted for 20 (33) minutes as indicated by the solid (dashed) magenta lines in Fig. 2. The ion energy–time spectrogram (Fig. 2a) shows significant fluxes in the energy range above about 5 keV. In the energy range of 1–5 keV, we intermittently observe an ion population mirrored in the Earth's magnetic field, as inferred

from numerical simulations¹⁰. These quasi-isotropic hot ions and middle-energy bouncing ions (Fig. 3a) are typical of the near-Earth central plasma sheet. The bulk velocity of the hot component was small ($\sim 40 \text{ km s}^{-1} = \sim 8 \text{ eV}$ in energy) through the event. These features show that GEOTAIL was located in the central plasma sheet throughout the event and that the characteristics of the hot ions before and after umbra passage are almost the same.

We note that GEOTAIL detected a third ion population, that is, an intense cold-ion signature just above a low-energy cut-off of ion counts (Fig. 2a) but only while in the solar umbra. The existence of the cut-off indicates that the spacecraft experienced negative charging down to the cut-off energy and that the negative spacecraft potential attracted the ambient positively charged ions toward the detector. This artificial acceleration due to the negative charging increases the apparent density of the ions, and precise estimation of the spacecraft potential is necessary in order to calculate the original density of the ambient cold ions.

To estimate the spacecraft potential, we first determine the lowest energy bins having valid ion counts (black lines in Fig. 2b). The actual potential must lie between the energies indicated by the two thin lines. The spacecraft potential began to decline during the penumbra, then gradually decreased with time during umbra passage, and rapidly returned to near zero after the umbra. We modelled the spacecraft potential during the eclipse with a smooth function described by a combination of polynomials as shown with the green line (Fig. 2b).

Once the spacecraft potential is determined, we can reconstruct the original distribution function of ions before the artificial acceleration due to the negative charging. The reconstructed phase space density during the umbra passage (Fig. 3b) shows that there exists a cold ion population that is below detection

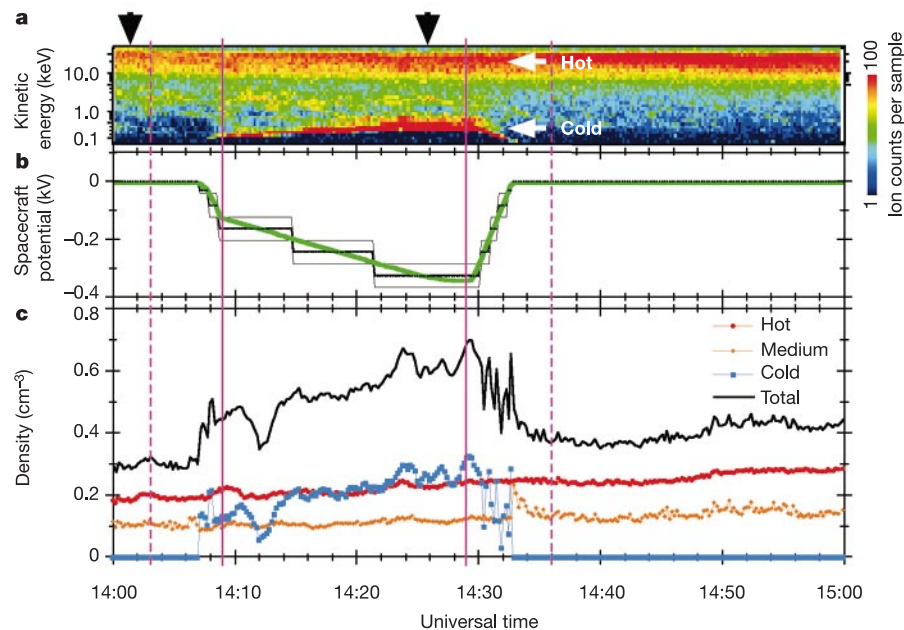


Figure 2 Ion observations by the GEOTAIL low-energy particle (LEP) instrument during the solar umbra event on 19 July 1997. The spacecraft position was $(X, Y, Z) \approx (-9.0, 0.2, 0.9)$ in GSM (geocentric solar magnetic) coordinates. **a**, An omni-directional energy–time spectrogram of ions, in which time series of ion counts at each kinetic energy in the spacecraft frame is colour-coded. The vertical solid (dashed) magenta lines indicate start and end times of the solar umbra (penumbra). **b**, Black lines display the energy range (thin lines) and centre energy (thick line) of the data bins just above the lower-energy cut-off of ion counts, and the green line shows the spacecraft potential estimated from the cut-off energy during the umbra or from single-probe measurements in sunlight. The potential in sunlight was around 0.004 kV and the potential gauge (potential difference

between ambient plasma and 0 kV potential level) is less than 0.002 kV. **c**, The density of hot (red), medium-energy (orange), and cold (blue) ion populations together with total ion density (black) calculated after the spacecraft potential correction. Because the instrument cannot distinguish ion species, these densities are calculated on the assumption that all ions are protons. If the cold-ion population consists of O^+ ions, the density becomes four times larger than the displayed value. The estimated densities of the hot, medium-energy and cold components have average uncertainties of 22%, 38% and 36%, respectively. Black arrows at the top correspond to the time when the distribution functions shown in Fig. 3 are observed.

threshold before the umbra (green dotted line). That is, the artificial acceleration due to charging in eclipse in effect lowered the instrument threshold and rendered these cold ions detectable—ordinarily they are ‘invisible’ in sunlight. Integrating reconstructed three-dimensional phase space densities separately over the two energy ranges corresponding to cold and middle ion populations, we can calculate the density of the two ion components. To calculate the density of the thermalized hot-ion population, we fitted a shifted-maxwellian distribution, which describes the thermal equilibrium of the plasma, to the corrected phase space density data above 5 keV (Fig. 3).

As shown in Fig. 2c, the densities of the hot (red) and medium-energy (orange) populations are almost constant throughout the interval from 14:00 to 15:00 UT. A remarkable point here is that the cold ion population (blue) during the umbra has a density comparable to that of the hot ions, and the density is also nearly constant in time when detected. The boundary conditions of the magnetosphere, that is, solar wind parameters, did not show a significant change throughout the event. In addition, similar values of the

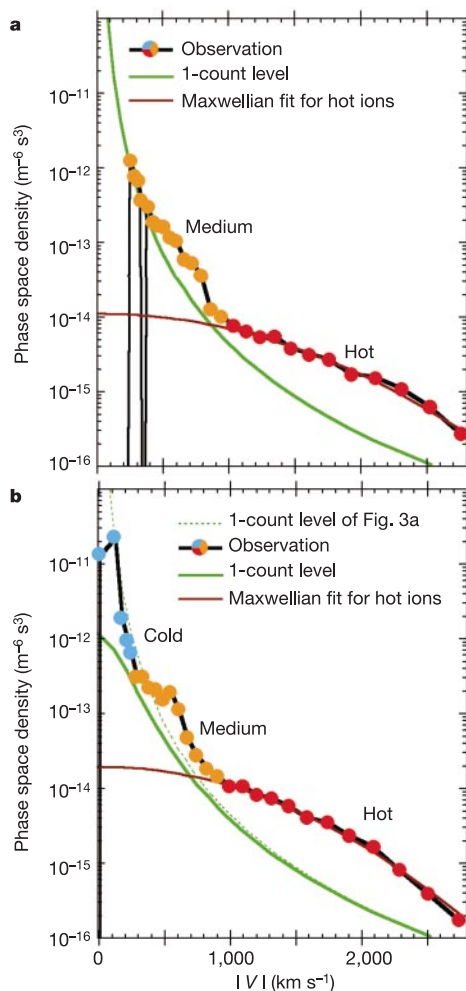


Figure 3 Examples of ion distribution functions observed before and during the solar umbra. **a**, Before the solar umbra. All ions are assumed to be protons and the phase space density is averaged over angle bins with valid counts. Red, orange and blue points correspond to energy ranges used to calculate densities of hot, medium-energy and cold ion populations, respectively. Results of a Maxwellian function fit for the hot-ion population are shown with brown lines. Green solid lines indicate the one-count level of the LEP instrument after spacecraft potential correction. The ratio of phase space density to the one-count level corresponds to counts per sampling time and shows the counting statistics of the data. **b**, During the solar umbra. The one-count level outside the umbra is also displayed with a dotted green line for reference. V represents ion velocity.

total electron density can be estimated from the observed cut-off frequency of continuum radiation in plasma wave data, before, during and after umbra passage. From these observational facts, we can deduce that the cold ions in the plasma sheet were present not only during the eclipse but also before and after the umbra. We note that this is not the only event that shows existence of cold ions in the plasma sheet. Of the 97 GEOTAIL solar umbra events in 1995–1998, 24 events at the anti-sunward distances between 9–18 Earth radii show the presence of cold ions in the hot plasma sheet (Fig. 1b).

Our current understanding of large-scale phenomena in circum-Earth space relies on the concept of magnetospheric convection¹⁴. The discovery of cold ions in the central plasma sheet sheds new light on magnetospheric convection as well as heating mechanisms for collisionless plasmas. In the presence of this convective plasma flow, low-energy ions from the lobe are always energized upon entry into the plasma sheet, according to simulations⁹. On the other hand, the existence of cold ions in the plasma sheet implies that a significant amount of plasma is somehow exempted from heating and acceleration in boundary layers upon entry into the plasma sheet. To avoid such a mysterious selective heating, one might think that the cold ions detected in the plasma sheet have not encountered the boundary heating region, but instead have flowed out from the ionosphere, gradually filling a plasma-sheet flux tube after the flux tube passed through the heating region (field line 5 in Fig. 1a). In that case, the ion outflow flux from the ionosphere¹⁵ predicts that this gradual filling will take several hours at least to supply the observed cold-ion density of 0.2 cm^{-3} . This is much longer than the transport time of the flux tube in ordinary magnetospheric convection according to current views¹⁴. Thus the observed density of around 0.2 cm^{-3} cannot easily be explained within the existing framework of space physics, unless convection is actually a much slower process, requiring on average more than several hours to go from the neutral line to 10 Earth radii. This idea of slow magnetospheric convection, even during multiple substorms, is at odds with conventional views of magnetospheric dynamics and requires us to reconsider fundamental processes of mass and energy transfer in the magnetosphere instead of requiring new mechanisms for the mysterious selective heating discussed above.

A cold plasma population with density comparable to that of the hot population carries a significant fraction of the mass density, and the effects of its inertia will not be negligible when considering the dynamics of large-scale phenomena in the Earth's magnetosphere. Although the instrument used here cannot directly provide ion composition information, comparison with electron density derived from wave data indicates that the upper limit of the O^+/H^+ density ratio is 0.5. If half of this high-density cold population consists of O^+ ions, it suggests an ionospheric source and may explain a large gap in the net outflow flux of terrestrial oxygen ions at high and low altitudes¹⁶. On the other hand, if the cold population consists of H^+ , the source can be either the ionosphere or the solar wind; the low temperature of the population makes the ionosphere source more likely. In either case, we need to determine how these ions can remain cold in the plasma sheet, exempted from the energization. Future multi-spacecraft missions that will enable the simultaneous observation of lobes, the plasma-sheet boundary layer, and the central plasma sheet will help us to discover what controls the acceleration and heating of collisionless plasma. □

Received 9 September 2002; accepted 18 February 2003; doi:10.1038/nature01502.

1. Priest, E. R. *et al.* Nature of the heating mechanism for the diffuse solar corona. *Nature* **393**, 545–547 (1998).
2. Scudder, J. D. On the causes of temperature change in inhomogeneous low-density astrophysical plasmas. *Astrophys. J.* **398**, 299–318 (1992).
3. Moses, R. W., Finn, J. M. & Ling, K. M. Plasma heating by collisionless magnetic reconnection: Analysis and computation. *J. Geophys. Res.* **98**, 4013–4040 (1993).
4. Ono, Y., Inomoto, M., Okazaki, T. & Ueda, Y. Experimental investigation of three-component magnetic reconnection by use of merging spheromaks and tokamaks. *Phys. Plasmas* **4**, 1953–1963 (2001).

5. Kodama, R. *et al.* Fast heating of ultrahigh-density plasma as a step towards laser fusion ignition. *Nature* **412**, 798–802 (2001).
6. Hoshino, M., Mukai, T., Yamamoto, T. & Kokubun, S. Ion dynamics in magnetic reconnection: Comparison between numerical simulation and Geotail observations. *J. Geophys. Res.* **103**, 4509–4530 (1998).
7. Baumjohann, W., Paschmann, G. & Cattell, C. A. Average plasma properties in the central plasma sheet. *J. Geophys. Res.* **94**, 6597–6606 (1989).
8. Gary, S. P. Electromagnetic ion/ion instabilities and their consequences in space plasmas—A review. *Space Sci. Rev.* **56**, 373–415 (1991).
9. Delcourt, D. C., Moore, T. E. & Chappell, C. R. Contribution of low-energy ionospheric protons to the plasma sheet. *J. Geophys. Res.* **99**, 5681–5689 (1994).
10. Arzner, K. & Scholer, M. Kinetic structure of the post plasmoid plasma sheet during magnetotail reconnection. *J. Geophys. Res.* **106**, 3827–3844 (2001).
11. Lakhina, G. S., Tsurutani, B. T., Kojima, H. & Matsumoto, H. “Broadband” plasma waves in the boundary layers. *J. Geophys. Res.* **105**, 27791–27831 (2000).
12. Saito, Y. *et al.* Slow-mode shocks in the magnetotail. *J. Geophys. Res.* **100**, 23567–23581 (1995).
13. Hirahara, M. *et al.* Acceleration and heating of cold ion-beams in the plasma sheet boundary layer observed with GEOTAIL. *Geophys. Res. Lett.* **21**, 3003–3006 (1994).
14. Nishida, A., Mukai, T., Yamamoto, T., Kokubun, S. & Maezawa, K. A unified model of the magnetotail convection in geomagnetically quiet and active times. *J. Geophys. Res.* **103**, 4409–4418 (1998).
15. Abe, T. *et al.* Observations of polar wind and thermal ion outflow by Akebono/SMS. *J. Geomagn. Geoelectr.* **48**, 319–325 (1996).
16. Seki, K. *et al.* On atmospheric loss of oxygen ions from earth through magnetospheric processes. *Science* **291**, 1939–1941 (2001).

Acknowledgements We thank all GEOTAIL science members for their collaboration.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.S. (e-mail: seki@stelab.nagoya-u.ac.jp).

Relating atomic-scale electronic phenomena to wave-like quasiparticle states in superconducting $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$

K. McElroy^{*†}, R. W. Simmonds^{*}, J. E. Hoffman^{*}, D.-H. Lee^{*‡}, J. Orenstein^{*†}, H. Eisaki[§], S. Uchida^{||} & J. C. Davis^{*†¶}

^{*} Department of Physics, University of California, and [†] Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[‡] Center for Advanced Study, Tsinghua University, Beijing 100084, China

[§] AIST, 1-1-1 Central 2, Umezono, Tsukuba, Ibaraki, 305-8568 Japan

^{||} Department of Physics, University of Tokyo, Yayoi, 2-11-16 Bunkyo-ku, Tokyo 113-8656, Japan

[¶] Department of Physics, LASSP, Cornell University, Ithaca, New York 14850, USA

The electronic structure of simple crystalline solids can be completely described in terms either of local quantum states in real space (*r*-space), or of wave-like states defined in momentum-space (*k*-space). However, in the copper oxide superconductors, neither of these descriptions alone may be sufficient. Indeed, comparisons between *r*-space^{1–5} and *k*-space^{6–13} studies of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (Bi-2212) reveal numerous unexplained phenomena and apparent contradictions. Here, to explore these issues, we report Fourier transform studies of atomic-scale spatial modulations in the Bi-2212 density of states. When analysed as arising from quasiparticle interference^{14–16}, the modulations yield elements of the Fermi-surface and energy gap in agreement with photoemission experiments^{12,13}. The consistency of numerous sets of dispersing modulations with the quasiparticle interference model shows that no additional order parameter is required. We also explore the momentum-

space structure of the unoccupied states that are inaccessible to photoemission, and find strong similarities to the structure of the occupied states. The copper oxide quasiparticles therefore apparently exhibit particle–hole mixing similar to that of conventional superconductors. Near the energy gap maximum, the modulations become intense, commensurate with the crystal, and bounded by nanometre-scale domains⁴. Scattering of the antinodal quasiparticles is therefore strongly influenced by nanometre-scale disorder.

Among the unexplained phenomena of Bi-2212 quasiparticles are: (1) the appearance of anti-nodal quasiparticles only below T_c (ref. 17); (2) the proportionality between the quasiparticle-peak intensity and both the superfluid and hole densities^{18,19}; (3) the ‘kink’ in the nodal-quasiparticle dispersion^{20,21}; (4) the apparent contradiction between nanoscale electronic disorder^{1–4} and quasiparticles well-defined in *k*-space²²; and (5) the nature of the vortex-core-induced electronic states²³. Here we focus on the possibility that some of these phenomena emerge from the special relationship between *r*-space and *k*-space that is characteristic of the strongly correlated electronic structure of the copper oxides.

To explore this relationship we use Fourier transform scanning tunnelling spectroscopy (FT-STs). In this technique, the tip-sample differential tunnelling conductance ($g = dI/dV$) is spatially mapped

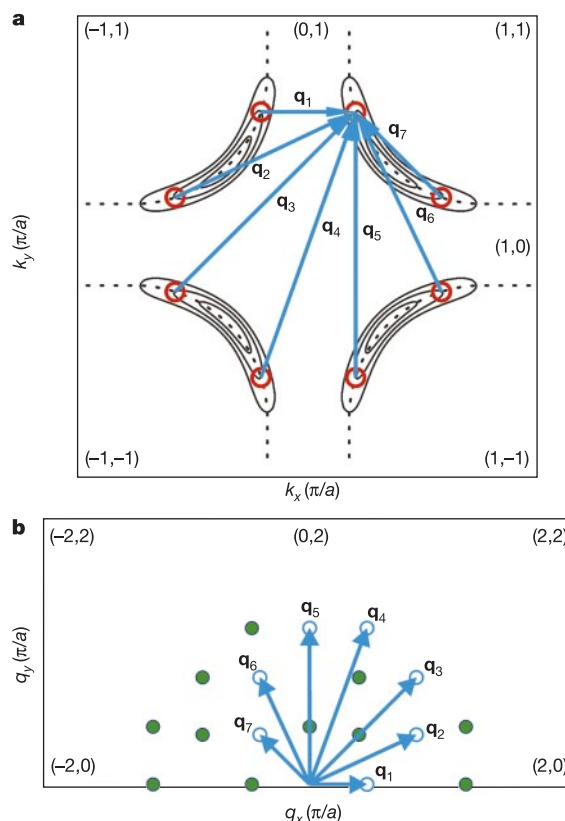


Figure 1 The expected wavevectors of quasiparticle interference patterns in a superconductor with electronic band structure like that of Bi-2212. **a**, Solid lines indicate the *k*-space locations of several banana-shaped quasiparticle CCE as they increase in size with increasing energy. As an example, at a specific energy, the octet of regions of high $|\nabla_k E(\mathbf{k})|^{-1}$ are shown as red circles. The seven primary scattering *q*-vectors interconnecting elements of the octet are shown in blue. **b**, Each individual scattering *q*-vector from this set of seven is shown as a blue arrow originating from the origin in *q*-space, and ending at a point given by a blue circle. The end points of all other inequivalent *q*-vectors of the octet model (as determined by mirroring each of the original seven in the symmetry planes of the Brillouin zone) are shown as solid green circles. Thus, if the quasiparticle interference model is correct, there would be sixteen inequivalent local maxima in the inequivalent half of *q*-space detectable by FT-STs.

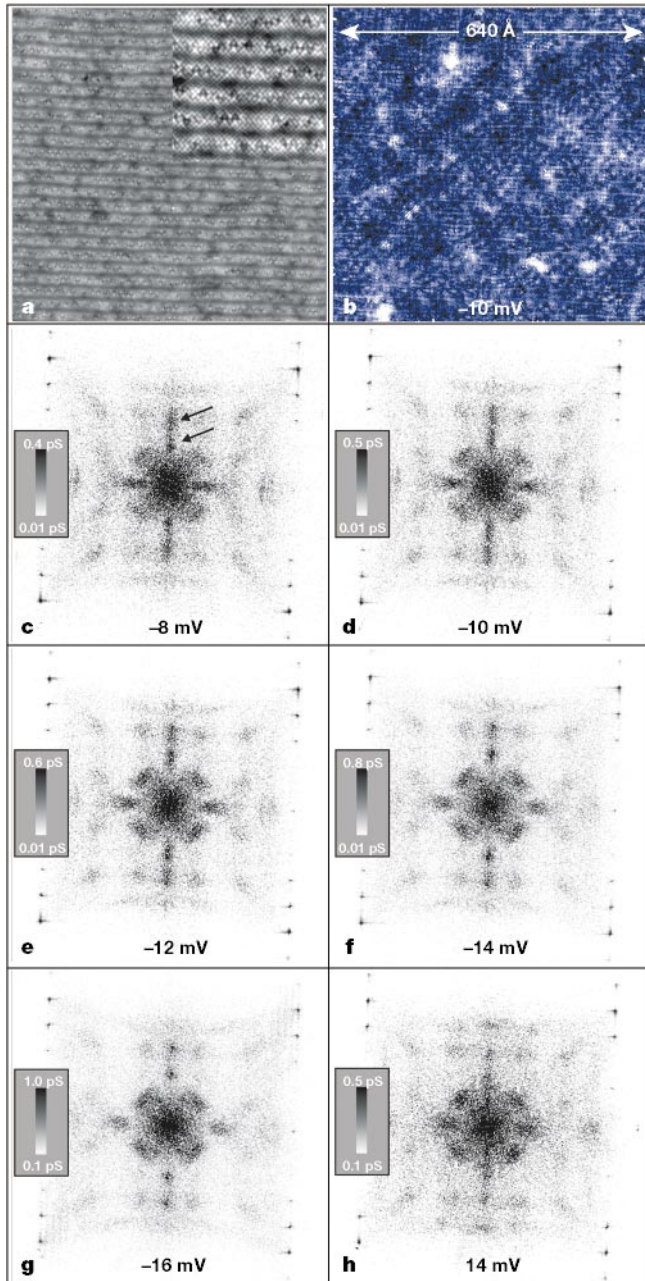


Figure 2 Atomic resolution images of the LDOS and the resulting Fourier-space images of the wavevectors making up the LDOS modulations. **a**, A topographic image of the BiO surface used, with location and resolution identical to the LDOS maps. The $\times 2$ magnification inset (from part of the same image) demonstrates the quality of atomic resolution achieved. **b**, Representative example of the real space $g(\mathbf{r}, \omega)$ in this field of view. All $g(\mathbf{r}, \omega)$ were acquired using the same atomic resolution and register. **c–g**, Examples of the $g(\mathbf{q}, \omega)$ at five different energies. The only non-dispersive signals (which are due to the supermodulation) are marked by arrows in **c**. The reciprocal atomic lattice is located at the square of intense points near the corners of each panel. One can readily see 12 of the 16 LDOS modulations of the quantum interference model. The dispersion and evolution of all the wavevectors of these modulations is evident in the differences between frames. Careful examination reveals that, in addition to the slowly dispersing $\mathbf{q}_1$ signal moving to $|\mathbf{q}| < \pi/2a_0$ with increasing energy, there is no additional signal above the noise at $\mathbf{q} = (1/4, 0)2\pi/a_0$ or $\mathbf{q} = (0, 1/4)2\pi/a_0$ (as proposed for a coexisting charge density wave (CDW) order parameter). **h**, The $g(\mathbf{q}, \omega)$ measured from the empty states at +14 meV above the Fermi level. It is similar but not identical to the $g(\mathbf{q}, \omega)$ at -14 meV in **f**, presumably because of the effects of normal-state band structure on the quasiparticle interference process.

at each bias voltage V . The result, $g(\mathbf{r}, \omega)$, is proportional to the local density of states (LDOS) at location $\mathbf{r}$ and energy $\omega = eV$. In metals²⁴ and more recently in a copper oxide superconductor⁵, quasiparticle interference has been observed. It is manifest as spatial modulations in $g(\mathbf{r}, \omega)$ whose wavelengths λ (or wavevectors $q = 2\pi/\lambda$) change with energy. In FT-STs, the $\mathbf{q}$ -vectors of these modulations can be determined from the locations of peaks in $g(\mathbf{q}, \omega)$, the Fourier transform magnitude of $g(\mathbf{r}, \omega)$.

We focus on Bi-2212, a material whose quasiparticle dispersion $E(\mathbf{k})$ has been mapped for the filled states by ARPES^{12,13}, and which exhibits four nodes in $\Delta(\mathbf{k})$. A contour plot of quasiparticle energy (ω) in $\mathbf{k}$ -space helps to visualize this situation. For $|\omega| < \Delta_0$ (the maximal gap), the contours of constant quasiparticle energy (CCE) are banana-shaped, as shown in Fig. 1a. With such unusual $\mathbf{k}$ -space structures, elastic quasiparticle scattering should produce striking consequences in $g(\mathbf{r}, \omega)$. The reason is that the quasiparticle density of states at ω , $n(\omega)$, is proportional to

$$\int_{E(\mathbf{k})=\omega} |\nabla_{\mathbf{k}} E(\mathbf{k})|^{-1} d\mathbf{k} \quad (1)$$

Because each 'banana' exhibits its largest rate of increase with energy $|\nabla_{\mathbf{k}} E(\mathbf{k})|^{-1}$ near its two ends, equation (1) shows that the primary contributions to $n(\omega)$ come from the octet of momentum-space regions centred around the end points $\mathbf{k}_j(\omega)$; $j = 1, 2, \dots, 8$. A particular octet is indicated by red circles in Fig. 1a.

During elastic scattering, a quasiparticle located near one element of the octet will very probably be scattered to the vicinity of another element, because of the large joint density of states between them. Owing to quantum interference⁵, such scattering produces spatial LDOS modulations with wavevector $\mathbf{q} = \mathbf{k}_{\text{final}} - \mathbf{k}_{\text{initial}}$ and with intensity proportional to $n_{\text{initial}}(\omega)n_{\text{final}}(\omega)$ (the joint density of states). Therefore pairs of states with high $n_{\text{initial}}(\omega)n_{\text{final}}(\omega)$ should generate the most intense LDOS modulations. The wavevectors $\mathbf{q}_i(\omega)$ of these modulations would then be determined by all possible pairs of points in the octet $\mathbf{k}_j(\omega)$. By considering just one $\mathbf{k}_j$ in a representative octet (Fig. 1a) we see that, at each energy, there are seven characteristic scattering wavevectors $\mathbf{q}_i(\omega)$; $i = 1, 2, \dots, 7$. These are indicated by blue arrows in the $\mathbf{k}$ -space of Fig. 1a and in the $\mathbf{q}$ -space of Fig. 1b. The model then predicts a total of $8 \times 7 = 56$ sets of scattering wavevectors. Only 32 of these are inequivalent and therefore 16 distinct $\pm\mathbf{q}$ pairs could be detected by FT-STs. Previously, no more than four sets of wavevectors have been detected^{5,25}.

We use float-zone-fabricated Bi-2212 single crystals which are as-grown (slightly overdoped) with $T_c = 86$ K. They are cleaved (at the BiO plane) in cryogenic ultrahigh vacuum and immediately inserted into the scanning tunnelling microscope head at 4.2 K. A search for all sets of $\mathbf{q}$ -vectors requires that each $g(\mathbf{r}, \omega)$ contain Fourier components with $|\mathbf{q}| \geq 2\pi/a_0$, so atomic resolution is required in all $g(\mathbf{r}, \omega)$. Even more importantly, to obtain sufficient $\mathbf{q}$ -space resolution $\Delta q = 2\pi/L$ so that the dispersions of all sets of $\mathbf{q}$ -vectors can be resolved, it is necessary that $g(\mathbf{r}, \omega)$ be measured in a field of view $L > 450$ Å. Fig. 2a shows the topographic image of the 640-Å-square field of view used for all studies reported here. We emphasize that this atomic resolution and position registry were maintained for all $g(\mathbf{r}, \omega)$ measurements. Figure 2b shows a typical example of a measured $g(\mathbf{r}, \omega)$ in which the LDOS modulations are clearly evident. Figure 2c–g are representative $g(\mathbf{q}, \omega)$ for several negative ω . No interpolation, smoothing or filtering is used on any image in this paper.

Neglecting effects of the supermodulation, all the $g(\mathbf{q}, \omega)$ in Fig. 2 are clearly fourfold symmetric and display numerous local maxima (dark regions) at different $\mathbf{q}$ for different energies. Quasiparticle interference has long been predicted for the copper oxides¹⁴. After ref. 5, new theoretical analyses^{15,16,26,27} were carried out, yielding predictions of up to 16 inequivalent peaks in $g(\mathbf{q}, \omega)$. These peaks are predicted to exhibit a specific fourfold symmetric pattern and to

disperse, self-consistently, along trajectories set by the Fermi surface, and $|\Delta(\mathbf{k})|$. The unprocessed data in Fig. 2 are in good qualitative agreement with these predictions.

To analyse these data further, we use the location of the peaks we can detect in $g(\mathbf{q}, \omega)$ between $-6 \text{ meV} > \omega > -30 \text{ meV}$ (from the quadrant or $\mathbf{q}$ -space without the supermodulation). This yields about 90 different $\mathbf{q}_i(\omega)$ whose magnitudes $|\mathbf{q}_i(\omega)|$ are plotted in

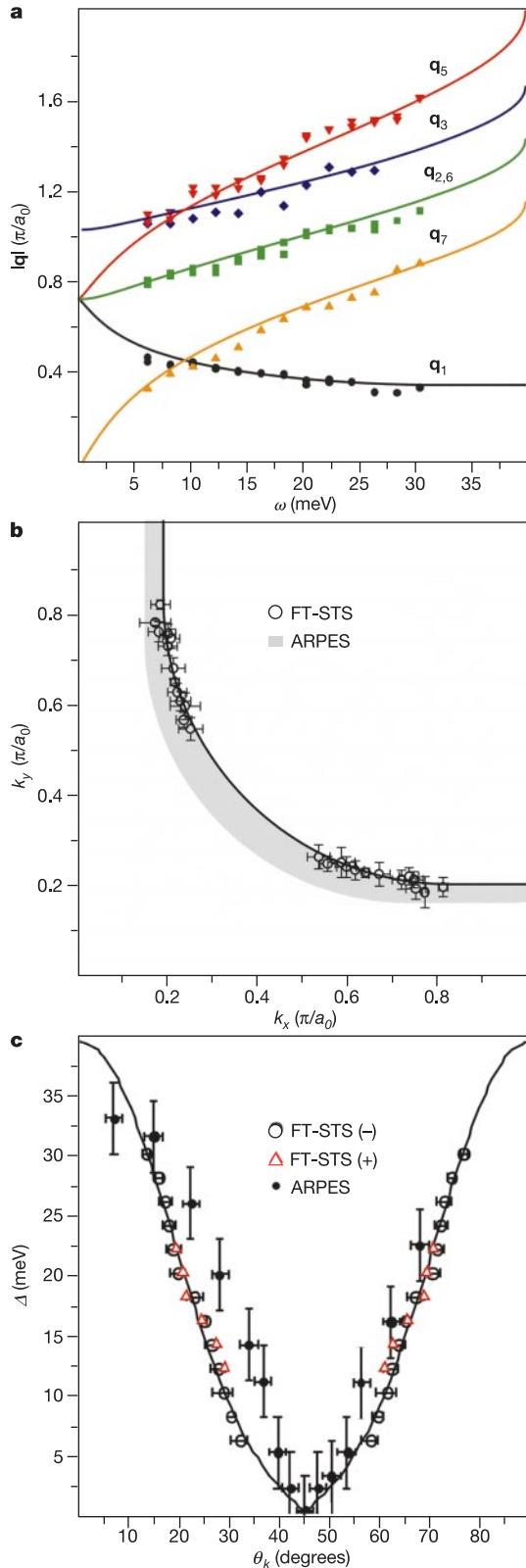


Fig. 3a. From these data, the locus from which the scattering originates $\mathbf{k}_s$ and energy-gap $\Delta(\mathbf{k})$ can be estimated by assuming that each:

$$\mathbf{q}_i(\omega) = \mathbf{k}_m(\omega) - \mathbf{k}_n(\omega) \quad (2)$$

where $\mathbf{k}_m(\omega)$ and $\mathbf{k}_n(\omega)$ are necessarily among the octet of states $\mathbf{k}_j(\omega)$ at ω . Using the mirror symmetries of the Brillouin zone plus equation (2), the location of the octet elements, $\mathbf{k}_s(\omega)$, can be determined from the $\mathbf{q}_i(\omega)$. Our FT-STs measured $\mathbf{k}_s(\omega)$ for the filled electronic states between $-6 \text{ meV} > \omega > -30 \text{ meV}$ is shown as open circles in Fig. 3b. We interpret it as the k -space trajectory of the ends of the CCE 'bananas' or, in quantum interference models^{14–16}, the normal-state Fermi surface. Because of the large number of $\mathbf{q}_i(\omega)$ measured for each ω , we can check the internal consistency of our data within the interference model using numerous independent observations. To do so we use at least five independent combinations of $\mathbf{q}_i(\omega)$ in equation (2) to identify $\mathbf{k}_s(\omega)$ at each ω . The error bars surrounding the data points in Fig. 3b are the statistical standard deviation in each $\mathbf{k}_s(\omega)$ resulting from these different combinations. Their small sizes demonstrate how these dispersive $\mathbf{q}$ -vectors are quite self-consistent within the interference model.

We can also measure the superconducting energy gap function $\Delta(\mathbf{k}_s)$ from the data in Fig. 3a. For a given $\omega = eV$, we find the associated $\mathbf{k}_s$ and plot $\omega(k_s)$, the energy at which quasiparticle scattering is intense along the trajectory $\mathbf{k}_s$. Within these models^{14–16}, this is the momentum dependence of the superconducting energy gap $\Delta(\mathbf{k}_s)$. Following ARPES notation, $\mathbf{k}_s$ is parameterized by using the angle θ_k about (π, π) . Our results for $\Delta(\theta_k)$ are shown as open circles in Fig. 3c and given by:

$$\Delta(\theta_k) = \Delta_0 [A \cos(2\theta_k) + B \cos(6\theta_k)] \quad (3)$$

with $\Delta_0 = 39.3 \text{ meV}$, $A = 0.818$, $B = 0.182$ (solid line in Fig. 3c).

We now discuss the various implications of our results. First, to compare $\mathbf{k}_s(\omega)$ and $\Delta(\theta_k)$ from FT-STs with those from photoemission, we plot the Fermi surface estimated from ARPES (on a sample of similar doping¹⁰) as the grey band in Fig. 3b and the ARPES-derived $\Delta(\theta_k)$ as solid circles in Fig. 3c. We note the good agreement between these results from two very different spectroscopic techniques. This agreement demonstrates the link between $\mathbf{r}$ -space and $\mathbf{k}$ -space characteristics of the copper oxide electronic structure, it gives enhanced confidence in both techniques because the matrix elements for photoemission and tunnelling are quite different, and it demonstrates that the proposed attribution of the nanoscale

Figure 3 The measured dispersion of all sets of $\mathbf{q}$ -vectors, the resulting FT-STs-derived locus of scattering, the anisotropic energy gap, and the relevant ARPES data for comparison. **a**, A plot of the magnitude of $\mathbf{q}_1$ through $\mathbf{q}_7$ (excluding $\mathbf{q}_4$) as a function of energy. The solid lines represent theoretical predictions based on fits to the $\mathbf{k}_s(\omega)$ and $\Delta(\mathbf{k}_s)$ as determined from FT-STs data. All the measured $\mathbf{q}$ -vectors are notably consistent with each other within the model. Below $\omega = 6 \text{ meV}$ LDOS modulations are difficult to detect, possibly because of strong effects of some unitary scattering resonances in the field of view or because of effects of the very low density of states near the nodal points¹⁵. **b**, The locus of scattering $\mathbf{k}_s(\omega)$ extracted using only the measured position of scattering vectors $\mathbf{q}_1$ to $\mathbf{q}_7$ (excluding $\mathbf{q}_4$). The solid line is a fit to the data, assuming the Fermi surface is the combination of a circular arc joined with two straight lines. The grey band represents measurements of the Fermi surface location made using the ARPES technique¹⁰. The error bars represent the statistical variations in a given $\mathbf{k}_s(\omega)$ when it is calculated using at least five different $\mathbf{q}_i(\omega)$. **c**, A plot of the energy gap $\Delta(\theta_k)$ determined from the filled-state measurements, shown as open circles. These were extracted using the measured position of scattering vectors $\mathbf{q}_1$ through $\mathbf{q}_7$ (excluding $\mathbf{q}_4$) in **a**. The solid line is a fit to the data. The filled circles represent $\Delta(\theta_k)$ determined using ARPES techniques¹⁰. The red, open triangles are the $\Delta(\theta_k)$ as determined from the unoccupied state measurements at positive bias as described in the text. They are in very good agreement with those from the filled states. The mean value of Δ_0 for this near-optimal sample was 39 meV .

electronic disorder detected by STS^{1–4} to surface damage not present in ARPES studies cannot be correct.

Second, the high-precision $g(\mathbf{q}, \omega)$ data presented here are relevant to proposals that LDOS modulations in Bi-2212 might result from the existence of a second charge-density wave order parameter with fixed $\mathbf{q}$ -vector, for example, stripes²⁵. In such models $g(\mathbf{q}, \omega)$ should exhibit only two non-dispersive peaks (or four non-dispersive peaks for twinned stripe domains). By contrast, quantum interference models predict 16 sets of dispersive $\mathbf{q}$ -vectors consistent with each other, the Fermi-surface, and $\Delta(\mathbf{k}_s)$. Clearly this is far more consistent with our data. More recent proposals^{26–29} suggest that a set of non-dispersive LDOS modulations due to fluctuations of a charge- or spin-ordered state might coexist with the quasiparticle interference patterns. We do not observe such a non-dispersive signal in addition to the quasiparticle interference effects, in the $g(\mathbf{q}, \omega)$ of these as-grown samples (except for crystalline effects). In principle, however, one cannot rule out the possibility of such a hypothetical non-dispersive signal because its intensity could be arbitrarily weak. Overall, our data demonstrate that quasiparticle interference is by far the predominant effect.

Third, we discuss the quantum mechanical description of the copper oxide quasiparticles. Bogoliubov quasiparticles are the excited states of a conventional Bardeen–Cooper–Schrieffer (BCS) superconductor: quantum-coherent mixtures of particles and holes. It is important to determine whether the copper oxide quasiparticles are of this type. A strong experimental indication consistent with Bogoliubov quasiparticles is the presence of two identical branches of quasiparticle dispersion $\pm E(\mathbf{k})$. Although efforts have been made to study the positive branch $+E(\mathbf{k})$ by

photoemission³⁰, it has proved challenging because the Fermi function terminates photoemission intensity from these unoccupied states. By using FT-STs we can probe the momentum-space structure of the positive branch by measuring $g(\mathbf{q}, \omega)$ at positive sample bias (tunnelling into unoccupied states). If identical positive and negative branches $\pm E(\mathbf{k})$ exist, the LDOS modulations at positive bias should be consistent with the negative-bias $\Delta(\mathbf{k})$ in Fig. 3. A representative example of positive bias $g(\mathbf{q}, \omega)$ measured at $\omega = +14$ meV is shown in Fig. 2h. Comparison with that at $\omega = -14$ meV shows them to be similar but not identical, as are all measured pairs $g(\mathbf{q}, \pm \omega)$. However, when the positive branch interference wavevectors $\mathbf{q}_i(+\omega)$ are measured, the deduced $\Delta(\mathbf{k})$ (triangles in Fig. 3c) is indistinguishable within errors from that of the filled-state $\Delta(\mathbf{k})$ (open circles in Fig. 3c). This provides evidence that, in momentum-space, the copper oxide quasiparticles are particle–hole superpositions, consistent with the Bogoliubov description.

Fourth, the data can also be used to explore implications of the nanoscale electronic disorder in Bi-2212 (refs 1–4). Figure 2 shows that at $|q| < 0.15$ there is a strong response in $g(\mathbf{q}, \omega)$ for all ω . These are apparent as dark regions near the centre of each panel in Fig. 2c–h. This may reflect long-wavelength inhomogeneity in the integrated LDOS¹ and, if so, reveals an obvious candidate for weak but ubiquitous potential scattering that could produce the LDOS modulations. We also note that the theoretical models, to date, are based on isolated point-like scatterers^{14,15,16} but the real scatterers are likely to be more complex in form. In that case, the character and strength of the scatterers could cause deviations of $k_s(\omega)$ (as determined here) from the real Fermi surface, whereas their spatial distribution could result in breaking the symmetry of $g(\mathbf{q}, \omega)$ under 90° rotations. New microscopic models and further experiments will be required to fully explore such effects.

A final new FT-STs observation relates to the antinodal quasiparticles which are at the heart of high- T_c superconductivity. Measurements of $g(\mathbf{r}, \omega)$ reveal intense LDOS modulations with wavevectors equal to the reciprocal lattice vectors $\mathbf{G}$ when $\omega \approx \Delta_0$ or equivalently when $\mathbf{k} \approx (\pi/a_0, 0)$. In a crystal, the electronic wavefunctions are a linear combination of states with wavevectors $\mathbf{k}$ and $\mathbf{k} + \mathbf{G}$ near the zone boundary. This mixing is due to Umklapp scattering off the crystal lattice and can produce intense LDOS modulations at $\mathbf{G}$ when $\mathbf{k} \approx (\pi/a_0, 0)$. However, as shown in Fig. 4, we unexpectedly find that for a given ω the Umklapp LDOS modulation signal is localized to the nanoscale regions where ω is equal to the local gap value. This implies strong nanoscale spatial variations in the quasiparticle dispersions near $\mathbf{k} = (\pi/a_0, 0)$ and therefore significant scattering. Thus, whatever the source of nanoscale electronic disorder, it appears to strongly influence the lifetimes of antinodal quasiparticles in Bi-2212. □

Received 30 July 2002; accepted 14 February 2003; doi:10.1038/nature01496.

- Pan, S. H. *et al.* Microscopic electronic inhomogeneity in the high- T_c superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. *Nature* **413**, 282–285 (2001).
- Cren, T., Roditchev, D., Sacks, W. & Klein, J. Nanometer scale mapping of the density of states in an inhomogeneous superconductor. *Euro. Phys. Lett.* **54**, 84–90 (2001).
- Howald, C., Fournier, P. & Kapitulnik, A. Inherent inhomogeneities in the tunneling spectra of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8-x}$ crystals in the superconducting state. *Phys. Rev. B* **64**, 100504 (2001).
- Lang, K. M. *et al.* Imaging the granular structure of high- T_c superconductivity in underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. *Nature* **415**, 412–416 (2002).
- Hoffman, J. E. *et al.* Imaging quasiparticle interference in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$. *Science* **297**, 1148–1151 (2002).
- Campuzano, J. C. *et al.* Fermi surfaces of $\text{YBa}_2\text{Cu}_3\text{O}_{6.9}$ as seen by angle-resolved photoemission. *Phys. Rev. Lett.* **64**, 2308–2311 (1990).
- Dessau, D. S. *et al.* Key features in the measured band structure of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$: Flat bands at E_F and Fermi surface nesting. *Phys. Rev. Lett.* **71**, 2781–2784 (1993).
- Aebi, P. *et al.* Complete Fermi surface mapping of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ (001): Coexistence of short range antiferromagnetic correlations and metallicity in the same phase. *Phys. Rev. Lett.* **72**, 2757–2760 (1994).
- Shen, Z.-X. *et al.* Anomalous large gap in the a-b plane of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. *Phys. Rev. Lett.* **70**, 1553–1556 (1993).
- Ding, H. *et al.* Angle-resolved photoemission spectroscopy study of the superconducting gap anisotropy in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. *Phys. Rev. B* **54**, R9678–R9681 (1996).

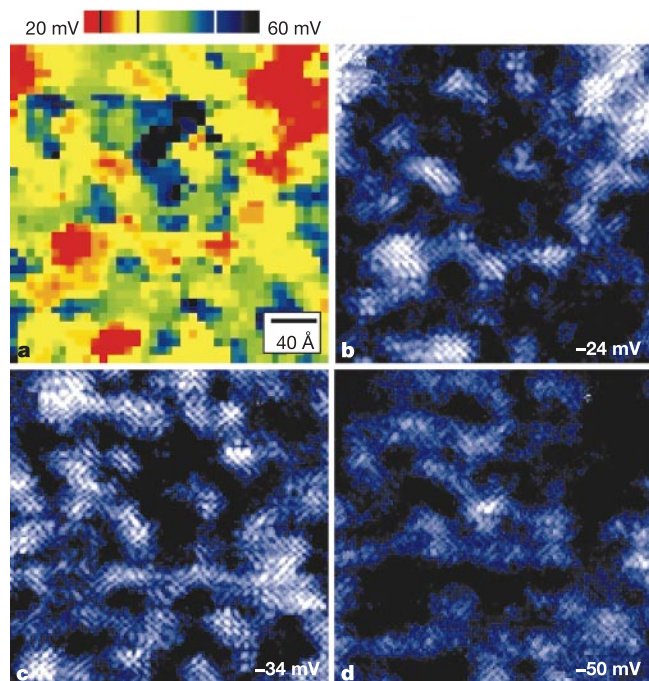


Figure 4 The electronic density of states modulations associated with antinodal quasiparticles at energies near the gap maximum. **a**, A map of the energy-gap magnitude⁴ in a particular area of the surface studied in this paper (see colour scale). **b–d**, $g(\mathbf{r}, \omega)$ measured at three energies, -24 meV, -34 meV and -50 meV respectively, in this exact field of view. One can immediately see, by comparison of panel **a** with the others, that wherever ω equals the local value of Δ , an intense ‘tweed’-like pattern exists $g(\mathbf{r}, \omega)$. The wavevectors of this pattern are the same in all three panels, either $\mathbf{q} = (2\pi/a_0, 0)$ or $\mathbf{q} = (0, 2\pi/a_0)$. Thus, LDOS modulations consistent with Umklapp scattering occur at different energies in adjacent nanoscale regions, signifying strong scattering of the antinodal quasiparticles.

11. Shen, Z.-X., Spicer, W. E., King, D. M., Dessau, D. S. & Wells, B. O. Photoemission studies of high- T_c superconductors: The superconducting gap. *Science* **267**, 343–350 (1995).
12. Damaselli, A., Shen, Z.-X. & Hussain, D. J. Angle-resolved photoemission spectroscopy of the cuprate superconductors. Preprint at (arXiv.org/cond-mat/0208504) (2002).
13. Campuzano, J. C., Norman, M. R. & Randeria, M. Photoemission in the high T_c superconductors. Preprint at (arXiv.org/cond-mat/0209476) (2002).
14. Byers, J. M., Flatté, M. E. & Scalapino, D. J. Influence of gap extrema on tunneling conductance near an impurity in an anisotropic superconductor. *Phys. Rev. Lett.* **71**, 3363 (1993).
15. Wang, Q. & Lee, D.-H. Quasiparticle scattering interference in high temperature superconductors. *Phys. Rev. B* **67**, 020511 (2003).
16. Zhang, D. & Ting, C. S. Energy-dependent LDOS modulation in cuprate superconductors. *Phys. Rev. B* **67**, 020511 (2003).
17. Fedorov, A. V. *et al.* Temperature dependent photoemission studies of optimally doped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **82**, 2179–2183 (1999).
18. Feng, D. L. *et al.* Signature of superfluid density in the single-particle excitation spectrum of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **289**, 277–281 (2000).
19. Dings, H. *et al.* Coherent quasiparticle weight and its connection to high- T_c superconductivity from angle-resolved photoemission. *Phys. Rev. Lett.* **87**, 227001 (2001).
20. Johnson, P. D. *et al.* Doping and temperature dependence of the mass enhancement observed in the cuprate $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **87**, 177007 (2001).
21. Lanzara, A. *et al.* Evidence for ubiquitous strong electron-phonon coupling in high-temperature superconductors. *Nature* **412**, 510–514 (2001).
22. Valla, T. *et al.* Evidence for quantum critical behavior in the optimally doped cuprate $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **285**, 2110–2113 (1999).
23. Hoffman, J. E. *et al.* A four unit cell periodic pattern of quasiparticle states surrounding vortex cores in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **295**, 466–469 (2002).
24. Crommie, M. F., Lutz, C. P. & Eigler, D. M. Imaging standing waves in a two-dimensional electron gas. *Nature* **363**, 524 (1993).
25. Howald, C., Eisaki, H., Kaneko, N. & Kapitulnik, A. Coexistence of charged stripes and superconductivity in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. Preprint at (arXiv.org/cond-mat/0201546) (2002).
26. Polkovnikov, A., Sachdev, S. & Vojta, M. Spin collective mode and quasiparticle contributions to STM spectra of d-wave superconductors with pinning. Preprint at (arXiv.org/cond-mat/0208334) (2002).
27. Podolsky, D. *et al.* Translational symmetry breaking in the superconducting state of the cuprates: Analysis of the quasiparticle density of states. Preprint at (arXiv.org/cond-mat/0204011) (2002).
28. Howald, C. *et al.* Periodic density of states modulations in superconducting $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. B* **67**, 014533 (2003).
29. Kivelson, S. A. *et al.* How to detect fluctuating order in the high-temperature superconductors. Preprint at (arXiv.org/cond-mat/0210683) (2002).
30. Campuzano, J. C. *et al.* Direct observation of particle-hole mixing in the superconducting state by angle-resolved photoemission. *Phys. Rev. B* **53**, R14737–R14740 (1996).

Acknowledgements We thank J. C. Campuzano, M. E. Flatté, P. Johnson, S. A. Kivelson, B. Lake, R. B. Laughlin, J. W. Loram, M. Norman, D. J. Scalapino, Z.-X. Shen, J. Tranquada and J. Zaanen for discussions and communications. This work was supported by an LDRD from the Lawrence Berkeley National Laboratory, the ONR, the NSF, and by Grant-in-Aid for Scientific Research on Priority Area (Japan), a COE grant from the Ministry of Education (Japan), and NEDO (Japan). J.E.H. is grateful for support from a Hertz Fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.C.D. (e-mail: jcdavis@ccmr.cornell.edu).

A subfemtotesla multichannel atomic magnetometer

I. K. Kominis^{*†}, T. W. Kornack^{*}, J. C. Allred[‡] & M. V. Romalis^{*}

^{*} Department of Physics, Princeton University, Princeton, New Jersey 08544, USA
[‡] Department of Physics, University of Washington, Seattle, Washington 98195, USA

The magnetic field is one of the most fundamental and ubiquitous physical observables, carrying information about all electromagnetic phenomena. For the past 30 years, superconducting quantum interference devices (SQUIDs) operating at 4 K have been unchallenged as ultrahigh-sensitivity magnetic field detectors¹, with a sensitivity reaching down to $1 \text{ fT Hz}^{-1/2}$ ($1 \text{ fT} = 10^{-15} \text{ T}$). They have enabled, for example, mapping of

the magnetic fields produced by the brain, and localization of the underlying electrical activity (magnetoencephalography). Atomic magnetometers, based on detection of Larmor spin precession of optically pumped atoms, have approached similar levels of sensitivity using large measurement volumes^{2,3}, but have much lower sensitivity in the more compact designs required for magnetic imaging applications⁴. Higher sensitivity and spatial resolution combined with non-cryogenic operation of atomic magnetometers would enable new applications, including the possibility of mapping non-invasively the cortical modules in the brain. Here we describe a new spin-exchange relaxation-free (SERF) atomic magnetometer, and demonstrate magnetic field sensitivity of $0.54 \text{ fT Hz}^{-1/2}$ with a measurement volume of only 0.3 cm^3 . Theoretical analysis shows that fundamental sensitivity limits of this device are below $0.01 \text{ fT Hz}^{-1/2}$. We also demonstrate simple multichannel operation of the magnetometer, and localization of magnetic field sources with a resolution of 2 mm.

Ultrasensitive magnetometers have found a wide range of applications, from condensed-matter experiments⁵ and gravitational wave detection⁶, to detection of NMR signals^{7,8}, studies of palaeomagnetism⁹, non-destructive testing¹⁰, and underwater ordinance detection¹¹. However, the most notable application of magnetic field sensors has been in the area of biomagnetism^{12,13}, that is, the detection of the weak magnetic fields produced by the human brain, heart and other organs. For example, measurements of the magnetic field produced by the brain are used to diagnose epilepsy, and to study neural responses to auditory and visual stimuli. Low-temperature superconducting quantum interference device (SQUID) sensors^{14–16}, which so far have dominated all of the above-mentioned applications, have reached sensitivity levels of $0.9\text{--}1.4 \text{ fT Hz}^{-1/2}$ with a pick-up coil area of the order of 1 cm^2 . In the low-frequency range of interest for biomagnetic studies ($<100 \text{ Hz}$) their noise is typically somewhat higher, whereas commercial SQUID magnetometers typically¹⁷ have noise of about $5 \text{ fT Hz}^{-1/2}$, partly due to magnetic noise generated by electrically conductive radiation-shielding of the liquid-helium dewars¹⁸.

Atomic magnetometers rely on a measurement of the Larmor precession of spin-polarized atoms in a magnetic field¹⁹. The fundamental, shot-noise-limited sensitivity of an atomic magnetometer is given by

$$\delta B = \frac{1}{\gamma \sqrt{n T_2 V t}} \quad (1)$$

where n is the number density of atoms, γ is their gyromagnetic ratio, T_2 is the transverse spin relaxation time, V is the measurement volume, and t is the measurement time²⁰. The value of γ in equation (1) depends on the details of the magnetometer operation. For a commonly used Zeeman transition with $\Delta m = 1$, $\gamma = g \mu_B / \hbar (2I + 1)$, where I is the nuclear spin of the alkali metal, μ_B is the Bohr magneton, and $g \approx 2$. In our magnetometer operating at zero field, the effective γ for sensitivity estimates is $\gamma = g \mu_B / \hbar$ (equation (7) of ref. 21).

Most atomic magnetometers use a polarized alkali-metal vapour (K, Rb, Cs), and their transverse spin relaxation time is limited by spin-exchange collisions between alkali atoms. In one implementation of such a magnetometer^{3,22}, the shot-noise sensitivity was estimated to be $0.3 \text{ fT Hz}^{-1/2}$ for a 500-cm^3 cell. In another state-of-the-art magnetometer², the actual sensitivity was estimated to be $1.8 \text{ fT Hz}^{-1/2}$ with a bandwidth of about 1 Hz and a measurement volume of $1,800 \text{ cm}^3$.

We recently demonstrated operation of a spin-exchange relaxation-free (SERF) magnetometer²¹ where broadening due to spin-exchange collisions is completely eliminated by operating at a high alkali-metal density in a very low magnetic field. The remaining broadening is determined by spin-relaxation collisions, which transfer spin angular momentum to rotational momentum of

[†] Present address: Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

atomic motion. They have much smaller cross-sections compared to spin exchange, yielding theoretical limits on magnetic field sensitivity below $0.01 \text{ fT Hz}^{-1/2}$ in 1 cm^3 . However, to realize such sensitivity in practical applications usually requires operating the device as a magnetic gradiometer to cancel common-mode magnetic field noise. Such operation is routinely used in SQUID detectors, but previous atomic gradiometers either used two identical magnetometers with a baseline of the order of 1 m (ref. 23) or had a relatively low magnetic field sensitivity of about $4 \text{ pT Hz}^{-1/2}$ (ref. 4). By adding ^4He buffer gas to the magnetometer cell we slow the diffusion of the K atoms, and demonstrate multichannel operation of the magnetometer with high sensitivity and adjacent channel spacing of only 3 mm . This allows us to cancel ambient magnetic field noise, and demonstrate magnetic source localization with high spatial resolution. In Fig. 1 we show the theoretical sensitivity of the magnetometer as a function of K density, including the effects of spin exchange and spin relaxation due to K–K and K–He collisions.

A diagram of the SERF magnetometer is shown in Fig. 2. Potassium atoms are spin-polarized by the pump laser along the $\hat{z}$ direction. We operate the magnetometer with all three components of the magnetic field close to zero. In this regime, a small B_y field rotates the spin polarization into the $\hat{x}$ direction (see the bottom inset in Fig. 2). The plane of the polarization of the probe beam is rotated in proportion to the $\hat{x}$ component of the spin polarization. Using a multichannel photodetector, we simultaneously measure the polarization of the probe beam and, therefore, the B_y field, at several adjacent points. It can be shown that the signal is not sensitive to small changes in B_x and B_z fields, and the device operates as a vector magnetometer. By taking a linear combination of the signals, we can make measurements of the first- and higher-order gradients of the magnetic field.

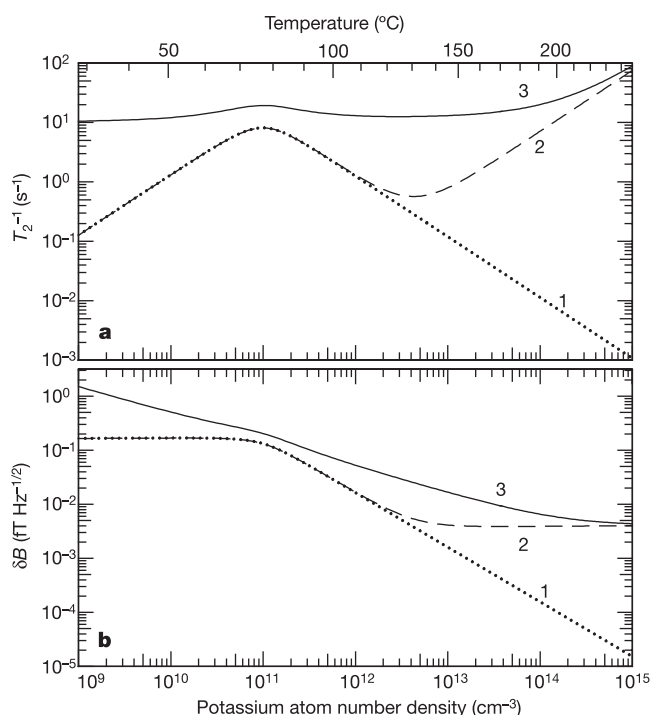


Figure 1 Relaxation rate and theoretical magnetic field sensitivity. **a**, Transverse relaxation rate T_2^{-1} due to alkali metal spin-exchange collisions²⁷ in a small magnetic field $B = 10 \mu\text{G}$ (case 1), spin-exchange and K–K spin-relaxation collisions (case 2), and the total transverse relaxation rate, also including the effect of K–He collisions for 2.9 atm of ^4He gas (case 3). **b**, Estimated magnetic field sensitivity shot-noise limit for cases 1–3 obtained from equation (1) with a measurement volume $V = 0.3 \text{ cm}^3$.

The magnetic field sensitivity data are shown in Fig. 3a. The magnetometer frequency response (Fig. 3b) is measured by applying a known oscillating B_y field at several frequencies. The frequency response depends on the optical-pumping and spin-relaxation rates²¹, and is well described by a single-pole, low-pass filter with a cut-off frequency of about 20 Hz . The noise in a single magnetometer channel of about $7 \text{ fT Hz}^{-1/2}$ (dashed line in Fig. 3a) is due to Johnson noise from our mu-metal shields with an inner diameter of 40 cm , and is consistent with estimates based on theoretical calculations¹⁸.

We form a first-order gradiometer by taking the difference between two adjacent magnetometer channels after correcting for a small difference in their absolute sensitivity. This procedure cancels the common magnetic field noise. Assuming that the remaining noise is uncorrelated, we divide the resulting noise level by $\sqrt{2}$ to obtain the intrinsic magnetic field sensitivity of each channel (solid line in Fig. 3a). Apart from a number of sharp peaks from technical sources of noise, the magnetic sensitivity is less than $1 \text{ fT Hz}^{-1/2}$ in the range $10\text{--}150 \text{ Hz}$, and averages to $0.54 \text{ fT Hz}^{-1/2}$ in the range $28\text{--}45 \text{ Hz}$. To our knowledge, this represents the best magnetic field sensitivity obtained to date in

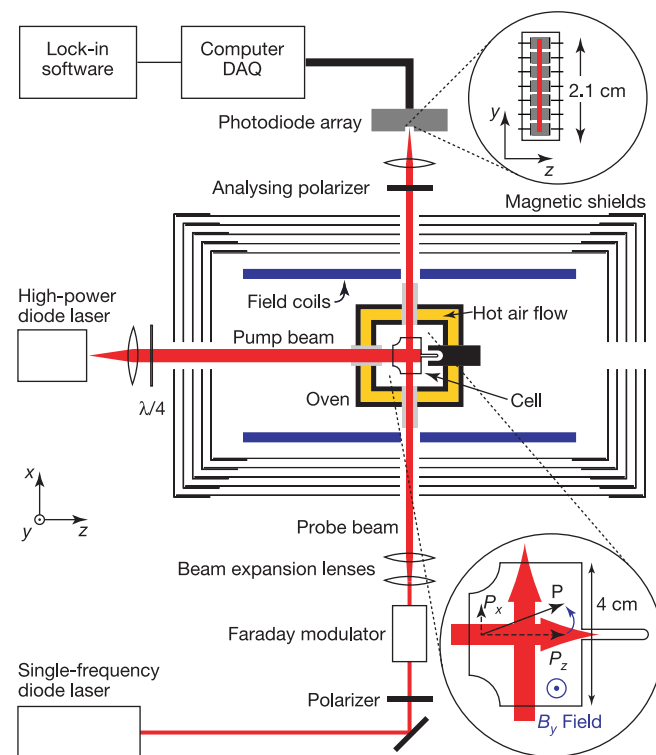


Figure 2 Experimental set-up. The diagram shows: magnetic shields with a shielding factor of 10^6 ; field coils producing calibrated, uniform fields along $\hat{x}$, $\hat{y}$ and $\hat{z}$ directions, and all five independent first-order field gradients; a T-shaped glass cell ($3 \times 4 \times 3 \text{ cm}$) with flat windows, containing a drop of K metal, 2.9 atm of ^4He and 60 torr of N_2 ; a double-wall oven heated to 180°C by flowing hot air to obtain a K atom number density of $n \approx 6 \times 10^{13} \text{ cm}^{-3}$; a circularly polarized 1 W broadband diode laser ('pump' laser) tuned to the centre of the D1 line at 770 nm ; a linearly polarized 100 mW single frequency laser ('probe' laser) detuned by 1 nm from the D1 resonance; a Faraday rotator modulating the plane of polarization of the probe laser with an amplitude $\alpha \approx 0.02 \text{ rad}$ at a frequency $f_{\text{mod}} = 2.9 \text{ kHz}$; beam-shaping optics that produce a collimated probe beam with a cross-section of $4 \text{ mm} \times 19 \text{ mm}$; a polarization analyser, orthogonal to the polarizer; a seven-element photodiode array (shown in the top inset), with element separation of 0.31 cm along the $\hat{y}$ -direction; and a 16-bit data acquisition system using a digital seven-channel lock-in amplifier to demodulate the signal proportional to the magnetic field B_y . Bottom inset, cross-section of the T-shaped cell, showing the rotation of the K polarization $\mathbf{P}$ into the $\hat{x}$ direction by an applied magnetic field B_y .

either superconducting or atomic magnetometers. The active measurement volume used by each channel is only 0.3 cm^3 .

The inset to Fig. 3b shows the magnetic noise in the difference between two channels as a function of the distance between them—that is, the baseline of the gradiometer. The probe beam slightly expands in the $\hat{y}$ direction, so the channel spacing is 0.28 cm , 10% smaller than the photodiode element separation. We expect the noise to increase with the baseline d of the gradiometer owing to the magnetic field gradient noise. A fit of the form $N = \sqrt{N_1^2 + N_2^2 + d^2 G^2}$, where N_1 and N_2 are the intrinsic noise levels in each channel, gives a magnetic field gradient noise $G = 1 \text{ fT cm}^{-1} \text{ Hz}^{-1/2}$, which is somewhat larger than our estimate of $0.5 \text{ fT cm}^{-1} \text{ Hz}^{-1/2}$ for the gradient noise produced by the magnetic shields¹⁸. This probably indicates that some noise comes from local sources—perhaps the metal in the temperature sensor near the cell. We can also form a second-order gradiometer using three adjacent channels. We find that the intrinsic sensitivity of each channel measured in this way is slightly better, but the improvement is not significant.

We also investigated the performance of the magnetometer in a multichannel imaging mode. We first apply a uniform oscillating magnetic field gradient dB_y/dy to check the linearity of the device (Fig. 4a). With the exception of the outer channels, which are not fully illuminated by the pump laser, the response is quite linear, and

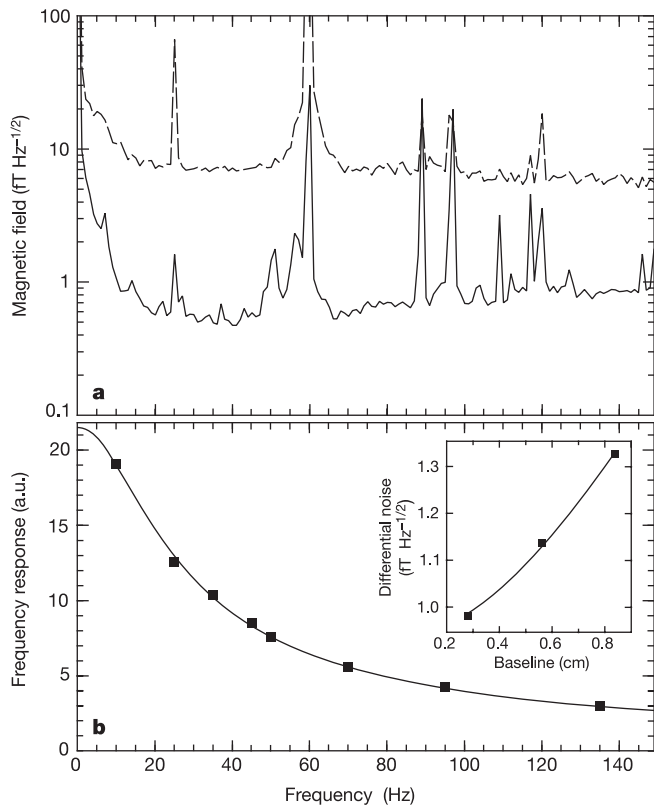


Figure 3 Magnetic field sensitivity and bandwidth of the magnetometer. Magnetic field noise in a single channel (**a**, dashed line), and intrinsic magnetic field sensitivity of a single channel extracted from the difference between adjacent channels (**a**, solid line). The magnetic field sensitivity data are obtained by recording the response of the magnetometer for about 100 s, performing a fast Fourier transform (FFT) without windowing; and calculating r.m.s. amplitudes in 1 Hz bins. A peak due to the calibrating B_y field is seen at 25 Hz. To obtain absolute field sensitivity, we divide the magnetometer FFT by a normalized frequency-response function shown in **b** with a fit to $A/(f^2 + B^2)^{1/2}$, where the bandwidth $B = 20 \text{ Hz}$. Inset, magnetic field noise in the channel difference as a function of the distance between channels. The fit assumes that the noise increases owing to contributions from magnetic field gradient noise. a.u., arbitrary units.

the measured gradient agrees to within 4% with the strength of the applied gradient. To simulate a biological source, we place a small coil about 5.3 cm from the centre of the magnetometer. We apply an oscillating current to the coil with a frequency of 25 Hz , and analyse the data in 1-s intervals (Fig. 4b). We fit the data to the magnetic dipolar field profile, and find that after 1 s of averaging the uncertainty in the distance to the dipole is 2 mm and the uncertainty in its absolute size is 13% .

The spatial resolution of the magnetic field measurements inside the magnetometer cell is limited by the diffusion of the K atoms. On the basis of a detailed model of diffusion, we determine the resolution to be about 2 mm for our conditions, slightly smaller than the spacing between the channels. The accuracy of localization of magnetic field sources outside the magnetometer depends on a number of factors, including the signal-to-noise ratio, the distance from the magnetometer, and the uniformity of the magnetometer response. Typically, the magnetic field sources can be localized to a fraction of the detector size, so localization uncertainty of the order of 0.2 mm can be expected in our magnetometer for sufficiently high signal-to-noise ratio.

In addition to its high sensitivity, the SERF magnetometer described here does not require cryogenic cooling, making it very attractive for a wide range of applications, particularly outside laboratory environments. The bandwidth and size of the magnetometer are well suited for detection of biological fields. Two technical modifications would be needed to use the magnetometer for magnetoencephalography (MEG). First, the magnetometer would have to be placed in larger magnetic shields (similar to the magnetically shielded rooms at present used for MEG with SQUID magnetometers²⁴), so as to accommodate a patient inside the shields. Second, more efficient thermal insulation and active cooling

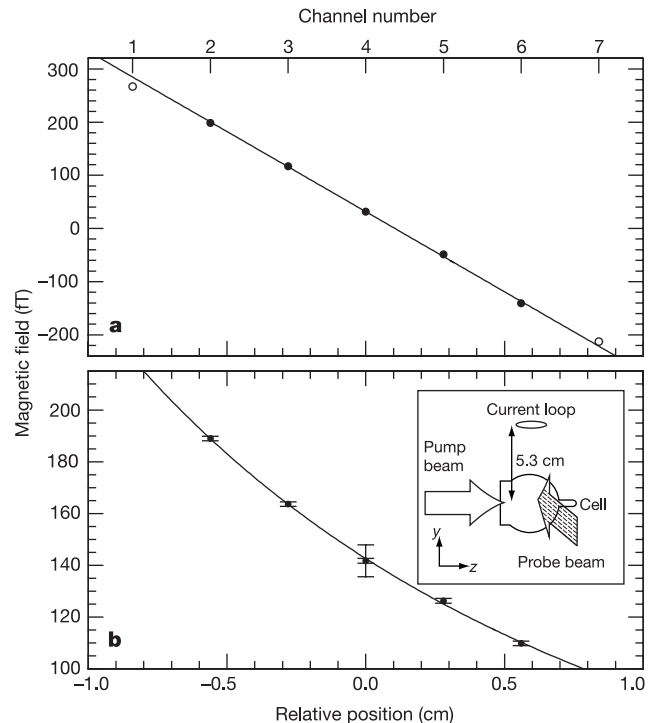


Figure 4 Magnetic gradient imaging. **a**, Measured response for an applied uniform gradient $\text{dB}_y/\text{dy} = 315 \text{ fT cm}^{-1}$ oscillating at 25 Hz . Only data represented by filled symbols are used in the linear fit, which gives a slope of 301 fT cm^{-1} . **b**, Measured response from a magnetic dipole $\mu = 1.25 \text{ uA cm}^{-2}$ located 5.3 cm away and oscillating at 25 Hz , with the magnetic dipole field fit. The large error bar on the middle data point represents the single-channel noise level after 1 s of integration. The small error bars represent the noise in the relative signal between adjacent channels.

would also be needed, to keep the outside surface of the sensor at room temperature.

Our magnetometer is in a very early stage of development, and many improvements are possible. For example, by using a two-dimensional photodiode array it should be possible to measure gradients and localize sources in two directions. The overall noise of the magnetometer could be further reduced by using superconducting shields, which do not have a Johnson noise component. In the absence of Johnson noise, simple averaging of the existing channels would yield a sensitivity of $0.2 \text{ fT Hz}^{-1/2}$. With more optimization, such as increased probe laser power and increased K atom density, it should be possible to approach the shot-noise-limited sensitivity in the range 10^{-2} – $10^{-3} \text{ fT Hz}^{-1/2}$. The thermal magnetic noise produced by the brain²⁵ is of the order of $0.1 \text{ fT Hz}^{-1/2}$, so an optimized version of this magnetometer should enable the maximum possible amount of information to be obtained about brain electrical activity. This may enable non-invasive studies of individual cortical modules in the brain²⁶, which have a size of 0.1–0.2 mm. □

Received 24 October 2002; accepted 4 February 2003; doi:10.1038/nature01484.

- Weinstock, H. (ed.) *SQUID Sensors: Fundamentals, Fabrication and Applications* (Kluwer Academic, Dordrecht, 1996).
- Aleksandrov, E. B. *et al.* Laser pumping in the scheme of an M_x -magnetometer. *Optics Spectrosc.* **78**, 292–298 (1995).
- Budker, D., Kimball, D. F., Rochester, S. M., Yashchuk, V. V. & Zolotarev, M. Sensitive magnetometry based on non-linear magneto-optical rotation. *Phys. Rev. A* **62**, 043403 (2000).
- Affolderbach, C., Stähler, M., Knappe, S. & Wynands, R. An all-optical, high sensitivity magnetic gradiometer. *Appl. Phys. B* **75**, 605–612 (2002).
- Tsuei, C. C. & Kirtley, J. R. Phase-sensitive evidence for d-wave pairing symmetry in electron-doped cuprate superconductors. *Phys. Rev. Lett.* **85**, 182–185 (2000).
- Harry, G. M., Jin, I., Paik, H. J., Stevenson, T. R. & Wellstood, F. C. Two-stage superconducting-quantum-interference-device amplifier in a high-Q gravitational wave transducer. *Appl. Phys. Lett.* **76**, 1446–1448 (2000).
- Greenberg, Ya. S. Application of superconducting quantum interference devices to nuclear magnetic resonance. *Rev. Mod. Phys.* **70**, 175–222 (1998).
- McDermott, R. *et al.* Liquid-state NMR and scalar couplings in microtesla magnetic fields. *Science* **295**, 2247–2249 (2002).
- Kirschvink, J. L., Maine, A. T. & Vali, H. Paleomagnetic evidence of a low-temperature origin of carbonate in the Martian meteorite ALH84001. *Science* **275**, 1629–1633 (1997).
- Tralshawala, N., Claycomb, J. R. & Miller, J. H. Practical SQUID instrument for non-destructive testing. *Appl. Phys. Lett.* **71**, 1573–1575 (1997).
- Clem, T. R. Superconducting magnetic gradiometers for underwater target detection. *Naval Engineers J.* **110**, 139–149 (1998).
- Hämäläinen, M. *et al.* Magnetoencephalography—theory, instrumentation, and applications to non-invasive studies of the working human brain. *Rev. Mod. Phys.* **65**, 413–497 (1993).
- Rodriguez, E. *et al.* Perception's shadow: long-distance synchronization of human brain activity. *Nature* **397**, 430–433 (1999).
- Zimmerman, J. E., Thiene, P. & Harding, J. T. Design and operation of stable RF-biased superconducting point-contact quantum devices, and a note on properties of perfectly clean metal contacts. *J. Appl. Phys.* **41**, 1572–1580 (1970).
- Drung, D., Bechstein, S., Franke, K. P., Scheiner, M. & Schurig, T. Improved direct-coupled dc SQUID read-out electronics with automatic bias voltage tuning. *IEEE Trans. Appl. Supercond.* **11**, 880–883 (2001).
- Oukhanski, N., Stolz, R., Zakosarenko, V. & Meyer, H. G. Low-drift broadband directly coupled dc SQUID read-out electronics. *Physica C* **368**, 166–170 (2002).
- Del Gratta, C., Pizzella, V., Tecchio, F. & Romani, G. L. Magnetoencephalography—a noninvasive brain imaging method with 1 ms time resolution. *Rep. Prog. Phys.* **64**, 1759–1814 (2001).
- Nononen, J., Montonen, J. & Katila, T. Thermal noise in biomagnetic measurements. *Rev. Sci. Instrum.* **67**, 2397–2405 (1996).
- Dupont-Roc, J., Haroche, S., Cohen-Tannoudji, C. Detection of very weak magnetic fields (10^{-9} gauss) by ^{87}Rb zero-field level crossing resonances. *Phys. Lett. A* **28**, 638–639 (1969).
- Budker, D. *et al.* Resonant nonlinear magneto-optical effects in atoms. *Rev. Mod. Phys.* **74**, 1153–1201 (2002).
- Allred, J. C., Lyman, R. N., Kornack, T. W. & Romalis, M. V. High-sensitivity atomic magnetometer unaffected by spin-exchange relaxation. *Phys. Rev. Lett.* **89**, 130801 (2002).
- Budker, D., Yashchuk, V. & Zolotarev, M. Nonlinear magneto-optic effects with ultra-narrow widths. *Phys. Rev. Lett.* **81**, 5788–5791 (1998).
- Aleksandrov, E. B., Balabas, M. V., Pasgalev, A. S., Vershovskii, A. K. & Yakobson, N. N. Double-resonance atomic magnetometers: From gas discharge to laser pumping. *Laser Phys.* **6**, 244–251 (1996).
- Kandori, A., Miyashita, T. & Tsukada, K. Cancellation technique of external noise inside a magnetically shielded room used for biomagnetic measurements. *Rev. Sci. Instrum.* **71**, 2184–2190 (2000).
- Varpula, T. & Poutanen, T. Magnetic field fluctuations arising from thermal motion of electric charge in conductors. *J. Appl. Phys.* **55**, 4015–4021 (1984).
- Ts'o, D. Y., Frostig, R. D., Lieke, E. E. & Grinvald, A. Functional organisation of primate visual cortex revealed by high-resolution optical imaging. *Science* **249**, 417–420 (1990).
- Happer, W. & Tam, A. C. Effect of rapid spin exchange on the magnetic-resonance spectrum of alkali vapors. *Phys. Rev. A* **16**, 1877–1891 (1977).

Acknowledgements This work was supported by the NIH, the Packard Foundation and Princeton University.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.R. (e-mail: romalis@princeton.edu).

Single-crystal gallium nitride nanotubes

Joshua Goldberger*, Rongrui He*, Yanfeng Zhang†, Sangkwon Lee†, Haoquan Yan*, Heon-Jin Choi† & Peidong Yang*†

* Department of Chemistry, University of California, Berkeley, California 94720, USA

† Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

Since the discovery of carbon nanotubes in 1991 (ref. 1), there have been significant research efforts to synthesize nanometre-scale tubular forms of various solids^{2–10}. The formation of tubular nanostructure generally requires a layered or anisotropic crystal structure^{2–4}. There are reports^{5,6,11} of nanotubes made from silica, alumina, silicon and metals that do not have a layered crystal structure; they are synthesized by using carbon nanotubes and porous membranes as templates, or by thin-film rolling. These nanotubes, however, are either amorphous, polycrystalline or exist only in ultrahigh vacuum⁸. The growth of single-crystal semiconductor hollow nanotubes would be advantageous in potential nanoscale electronics, optoelectronics and biochemical-sensing applications. Here we report an 'epitaxial casting' approach for the synthesis of single-crystal GaN nanotubes with inner diameters of 30–200 nm and wall thicknesses of 5–50 nm. Hexagonal ZnO nanowires were used as templates for the epitaxial overgrowth of thin GaN layers in a chemical vapour deposition system. The ZnO nanowire templates were subsequently removed by thermal reduction and evaporation, resulting in ordered arrays of GaN nanotubes on the substrates. This templating process should be applicable to many other semiconductor systems.

We grew arrays of ZnO nanowires on (110) sapphire wafers using a vapour deposition process developed in our laboratory¹². These ZnO nanowire arrays were placed inside a reaction tube for GaN chemical vapour deposition. Trimethylgallium and ammonia were used as precursors, and were fed into the system with argon or nitrogen carrier gas. The deposition temperature was set at 600–700 °C. After the GaN deposition, the samples were treated at 600 °C with 10% H₂ in argon to remove the ZnO nanowire templates.

Figure 1a shows a scanning electron microscopy (SEM) image of the starting ZnO nanowire array templates. These nanowires have uniform lengths of 2–5 μm and diameters of 30–200 nm. They are well faceted with hexagonal cross-sections (Fig. 1a inset), exhibiting {110} planes on the sides. After the GaN deposition and template removal, the colour of the sample turns from white to yellowish or darker. The morphology of the initial nanowire arrays was maintained (Fig. 1b), except for the increase in the diameters of the resulting nanostructures. The nanostructures appear less faceted than the starting ZnO nanowire template. Compositional analysis on the final product shows little Zn signal. X-ray diffraction (XRD) on the sample shows only (00l) diffraction peaks of the

wurtzite GaN structure (Fig. 1c), which indicates excellent epitaxy/texturing for the GaN coating.

The sample in Fig. 1b was dispersed onto a transmission electron microscopy (TEM) grid for further structural analysis. It was found that most of the nanostructures exhibit tubular structures with uniform wall thicknesses (Fig. 2a). The nanotubes have inner diameters ranging from 30 to 200 nm, similar to the ZnO nanowire arrays, and wall thicknesses between 5 and 50 nm (Fig. 2a–c). Most of the tubes have only one end open, but some tubes with both ends open were also observed. Figure 2a–c shows representative images of these GaN nanotubes. These observations are consistent with our SEM studies, where round and less-faceted ends are observed after the GaN coating (Fig. 1b). We thus conclude that those open nanotube ends were originally located at the GaN and substrate interface, and were broken open during TEM sample preparation. Indeed, we have frequently observed these open ends on the substrate surface together with the corresponding nanotubes (Fig. 1b inset). TEM studies also indicate that the inner cross-section of the nanotubes remains pseudo-hexagonal after the template removal (Supplementary Information).

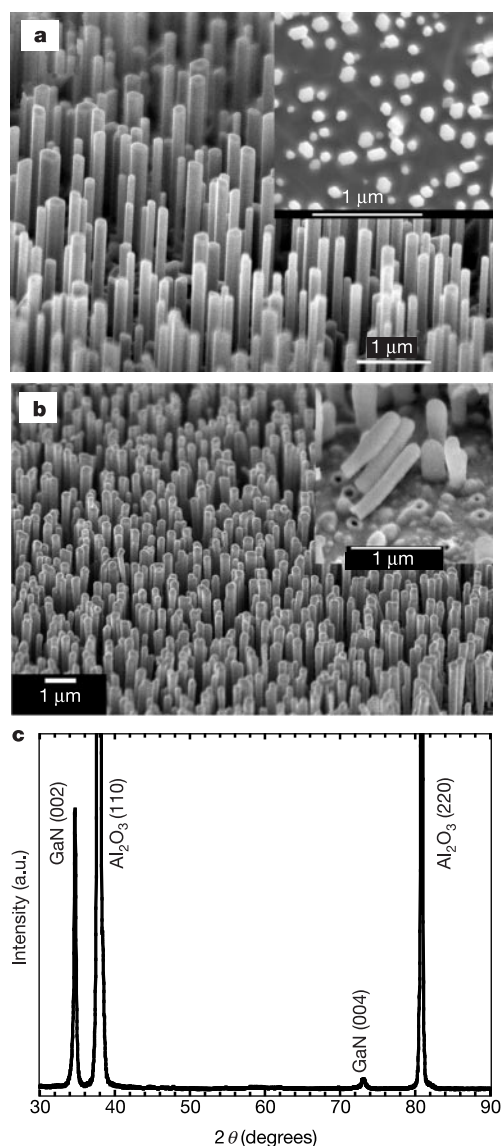


Figure 1 Arrays of ZnO nanowires and GaN nanotubes. Shown are SEM images of the ZnO nanowire template arrays (a), and the resulting GaN nanotube array (b). Inset in a shows cross-sections of the ZnO nanowires. Inset in b shows the fractured interface between the GaN nanotubes and the substrate. c, XRD of the GaN nanowire array.

Electron diffraction measurements on these GaN nanotubes indicates that they are single crystals. Figure 2e inset shows one such diffraction pattern taken along the $[110]$ zone axis. It can be readily seen that the nanotube is oriented along the c axis of the wurtzite GaN structure. This is consistent with the XRD data, where only (00l) peaks were observed. Along the tube axis, a lattice spacing of 0.51 nm for (001) planes of the wurtzite structure can be readily resolved on high-resolution TEM images of both the tube surface (Fig. 2d) and the inside of the tubes (Fig. 2e).

Compositional line profiles probed by energy dispersive X-ray spectroscopy (EDX) shows well-correlated gallium and nitrogen signals across the tube walls (Fig. 2f), indicating stoichiometric GaN formation during the deposition. This is also clearly reflected in the

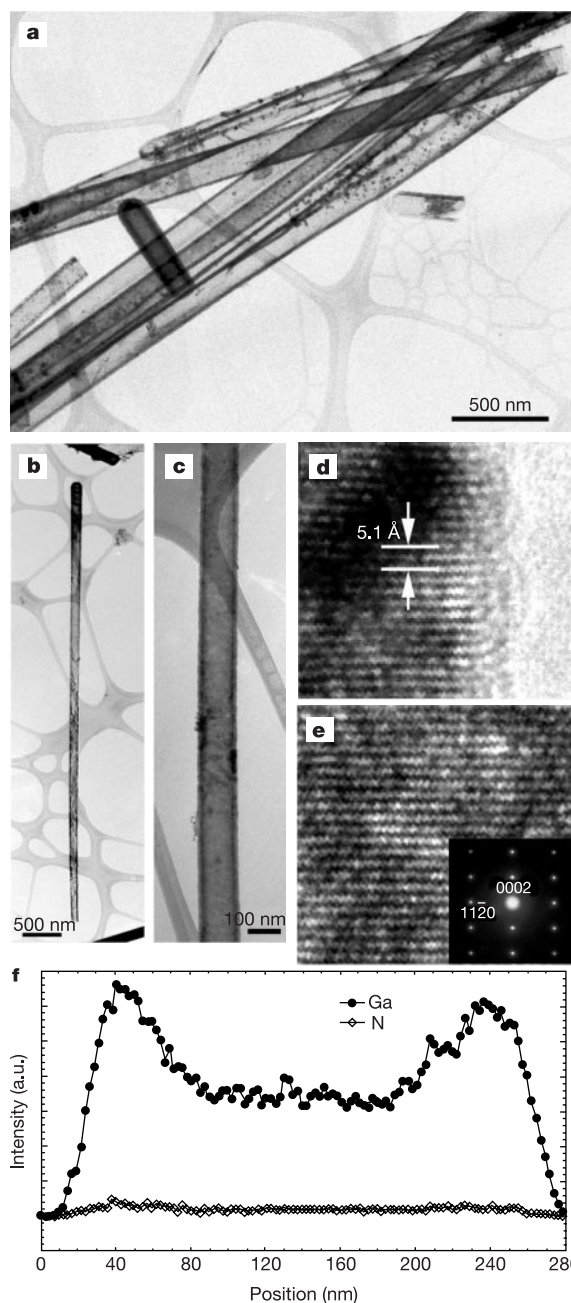


Figure 2 Structural characterization of GaN nanotubes. a–c, TEM images of the GaN nanotubes (a–c). d, High-resolution TEM images of a GaN nanotube wall at the surface, and e, the inside of the tube. Inset in e is an electron-diffraction pattern of the nanotube along the $[110]$ zone axis. f, Compositional line profile across the nanotube probed by EDX spectroscopy.

electron energy loss spectra recorded on these nanotubes, where strong nitrogen signals were observed (Supplementary Information). We note that the interfacial diffusion between the GaN layer and the ZnO nanowire templates results in a small amount of Zn or O incorporation within the GaN tube wall.

Taken together, the above measurements show that we have successfully prepared high-density arrays of single-crystal GaN nanotubes on sapphire substrates. We point out that the GaN nanotube formation reported here differs from all previous work on inorganic nanotubes^{1–6}. Most of the previous studies on inorganic nanotubes have been on materials with layered structures (for example, VO_x, MoS₂, NiCl₂, BN)^{1–4}. For those studies on materials that do not have structural anisotropy, templating approaches¹³ (in porous alumina) have been generally used, which result in predominantly amorphous or polycrystalline tubes⁵. In the process we report here, an ‘epitaxial casting’ approach was taken. Hexagonal-shaped single-crystal ZnO nanowires are used as templates initially during the GaN deposition (Fig. 3). As ZnO and GaN both have wurtzite crystal structures and have similar lattice constants (ZnO: $a = 3.249 \text{ \AA}$, $c = 5.207 \text{ \AA}$; GaN: $a = 3.189 \text{ \AA}$, $c = 5.185 \text{ \AA}$), GaN can grow epitaxially on the side {110} planes of these ZnO nanocylinders and form a thin single-crystal GaN layer. Once the ZnO nanocylinders were coated with a thin GaN sheath, they were subsequently removed by thermal processes. There are two possible mechanisms for the removal of ZnO templates. First, ZnO can be chemically etched by ammonia at high temperature¹⁴. Prolonged heating of samples after GaN coating in NH₃ readily yields pure GaN nanotubes. The other approach is to use thermal reduction at high temperatures (for example, 600 °C in H₂). The single-crystal wurtzite GaN nanotubes reported here differ fundamentally from theoretically simulated GaN nanotubes, where a metastable graphitic GaN structure was proposed¹⁵.

We have confirmed this ‘epitaxial casting’ mechanism by TEM studies. Arrays of GaN nanotubes with their ZnO nanowire templates partially removed are shown in Fig. 4a, b. Note that there is a thin layer of porous GaN film at the bottom of these nanotubes. In addition, the residue ZnO nanowire templates remain in the upper portion of the sealed GaN nanotubes. These two observations suggest that the zinc and oxygen species (generated during the thermal chemical etching process) escape from the GaN nanotubes primarily through the porous GaN layer (Supplementary Information). Electron diffraction (Fig. 4b insets) shows an identical set of diffraction pattern for both the tube and the core–sheath region, indicating the wurtzite GaN growth is epitaxial. The core–sheath nanostructure can be considered as a seamless single domain of a wurtzite GaN/ZnO structure type. Furthermore, comparison of EDX line profiles across the GaN nanotube (Fig. 4d) and the ZnO–GaN core–sheath structure (Fig. 4c) unambiguously supports the growth mechanism of GaN nanotube on the ZnO nanowire

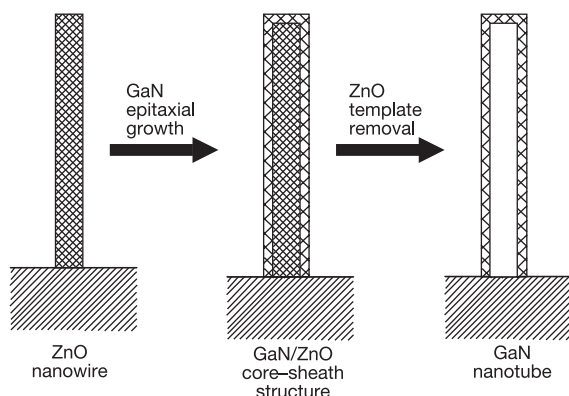


Figure 3 Schematic illustration of the ‘epitaxial casting’ process for making single-crystal GaN nanotubes.

templates. Once the ZnO nanocylinder is removed, single-crystal tubes of GaN result. To the best of our knowledge, this is the first example of the formation of single-crystal GaN nanotubes. (There was a previous report on GaN nanotube formation, but those tubes were highly irregular and polycrystalline¹⁶.) We note that microscale tubes of ZnO have been prepared in solution through a preferential chemical dissolution process¹⁷.

The electrical and optical characteristics of these single-crystal GaN nanotubes are comparable to those of high-quality GaN epilayers grown on ZnO substrates¹⁸, as well as those of GaN nanowires^{19,20}. We measured low-temperature photoluminescence spectra of our nanotubes using fourth-harmonic output of a YAG

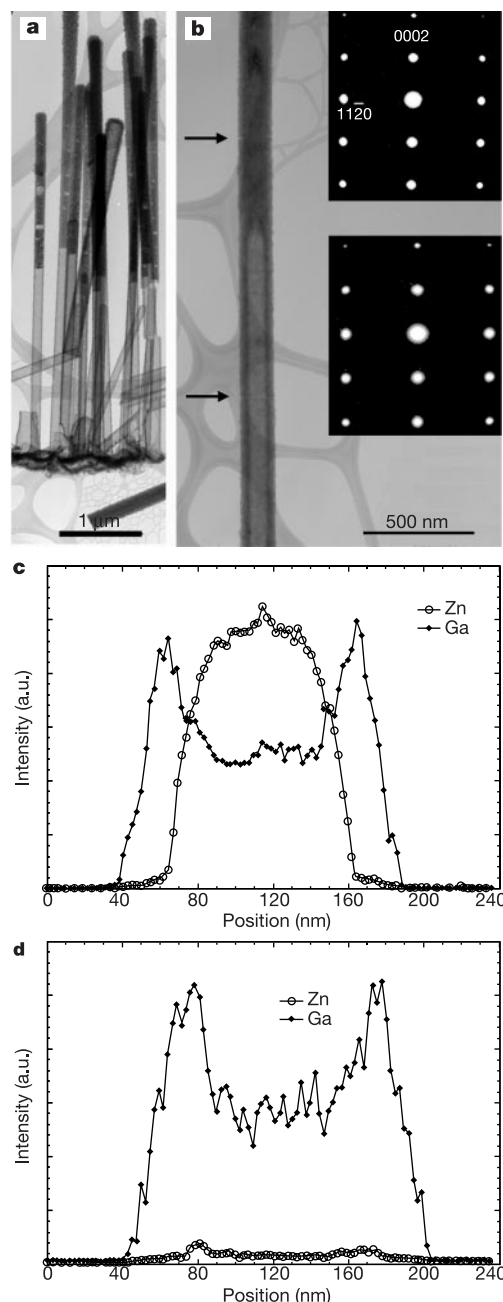


Figure 4 Characterization of the interface between the GaN tube and the ZnO wire. **a, b**, TEM images of arrays of GaN nanotubes with their ZnO nanowire templates partially removed. Insets in **b** are electron-diffraction patterns recorded on the core–sheath and the pure tube region along the $[110]$ zone axis. **c, d**, Compositional (Ga and Zn signals) line profiles for core–sheath GaN/ZnO (**c**) and the GaN tube region (**d**) indicated by upper and lower arrows in **b**, respectively.

laser (266 nm) as excitation source. No midgap yellow emission was observed. Band-edge emission was observed in these nanotube samples between 375 and 360 nm, with the thinner tubes emitting at shorter wavelengths (Supplementary Information). This slight blueshift of the emission¹⁸ could be attributed to the quantum confinement effect, as some of the nanotubes have walls as thin as 5 nm, which is smaller than the exciton Bohr radius of GaN. Electron transport measurements indicate the resistances of these nanotubes are of the order of 10 M Ω at room temperature, and increase with decreasing temperature (Supplementary Information), similar to those of high-quality GaN nanowires^{19,20}.

The semiconductor nanotubes that we report here are mechanically robust, and electrically and optically active. They could therefore offer opportunities for further fundamental research, as well as for technological applications in nanocapillary electrophoresis, nanofluidic biochemical sensing, nanoscale electronics and optoelectronics²¹. The successful preparation of single-crystal GaN nanotubes using this 'epitaxial casting' approach suggests that it is possible to prepare single-crystal nanotubes of other inorganic solids that have non-layered crystal structures^{22,23}. □

Received 14 November 2002; accepted 26 February 2003; doi:10.1038/nature01551.

1. Iijima, S. Helical microtubules of graphitic carbon. *Nature* **354**, 56–58 (1991).
2. Tenne, R. & Zettl, A. K. Nanotubes from inorganic materials. *Top. Appl. Phys.* **80**, 81–112 (2001).
3. Tenne, R. Inorganic nanoclusters with fullerene-like structure and nanotubes. *Prog. Inorg. Chem.* **50**, 269–315 (2001).
4. Patzke, G. R., Krumeich, F. & Nesper, R. Oxidic nanotubes and nanorods—Anisotropic modules for a future nanotechnology. *Angew. Chem. Int. Edn Engl.* **41**, 2446–2461 (2002).
5. Martin, C. R. Nanomaterials—a membrane-based synthetic approach. *Science* **266**, 1961–1965 (1994).
6. Ajayan, P. M. *et al.* Carbon nanotubes as removable templates for metal-oxide nanocomposites and nanostructures. *Nature* **375**, 564–566 (1996).
7. Yang, S. M. *et al.* Formation of hollow helicoids in mesoporous silica: Supramolecular origami. *Adv. Mater.* **11**, 1427–1430 (1999).
8. Kondo, Y. & Takanayagi, K. Synthesis and characterization of helical multi-shell gold nanowires. *Science* **289**, 606–608 (2000).
9. Li, Y. *et al.* Bismuth nanotubes. *J. Am. Chem. Soc.* **123**, 9904–9905 (2001).
10. Wu, Y. & Yang, P. Melting and welding semiconductor nanowires in nanotubes. *Adv. Mater.* **13**, 520–523 (2001).
11. Schmidt, O. G. & Eberl, K. Thin solid films roll up into nanotubes. *Nature* **410**, 168 (2001).
12. Huang, M. *et al.* Room-temperature ultraviolet nanowire nanolasers. *Science* **292**, 1897–1899 (2001).
13. Caruso, R. A. & Antonietti, M. Sol-gel nanocoating: an approach to the preparation of structured materials. *Chem. Mater.* **13**, 3272–3282 (2001).
14. Hamdani, F. *et al.* Effect of buffer layer and substrate surface polarity on the growth by molecular beam epitaxy of GaN on ZnO. *Appl. Phys. Lett.* **71**, 3111–3113 (1997).
15. Lee, S. M. *et al.* Stability and electronic structure of GaN nanotubes from density-functional calculations. *Phys. Rev. B* **60**, 7788–7791 (1999).
16. Li, J. Y. *et al.* Synthesis of GaN nanotubes. *J. Mater. Sci. Lett.* **20**, 1987–1988 (2001).
17. Vayssieres, L., Keis, K., Hagfeldt, A. & Lindquist, S. Three-dimensional array of highly oriented crystalline ZnO microtubes. *Chem. Mater.* **13**, 4395–4398 (2001).
18. Hamdani, F. *et al.* Microstructure and optical properties of epitaxial GaN on ZnO (0001) grown by reactive molecular beam epitaxy. *J. Appl. Phys.* **83**, 983–990 (1998).
19. Huang, Y., Duan, X., Cui, Y. & Lieber, C. M. Gallium nitride nanowire nanodevices. *Nano Lett.* **2**, 101–104 (2002).
20. Kim, J. *et al.* Electrical transport properties of individual gallium nitride nanowires synthesized by chemical vapor deposition. *Appl. Phys. Lett.* **80**, 3548–3550 (2002).
21. Schoening, M. & Poghosian, A. Recent advances in biologically sensitive field-effect transistors (BioFETs). *Analyst* **127**, 1137–1151 (2002).
22. Lauhon, L. J., Gudiksen, M. S., Wang, D. & Lieber, C. M. Epitaxial core-shell and core-multishell nanowire heterostructures. *Nature* **420**, 57–61 (2002).
23. He, R., Law, M., Fan, R., Kim, F. & Yang, P. Functional bimorph composite nanotapes. *Nano Lett.* **2**, 1109–1112 (2002).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank the National Center for Electron Microscopy for the use of their facilities. This work was supported by the Camille and Henry Dreyfus Foundation, the Research Corporation, the Hellman Family Faculty Foundation and the Beckman Foundation. J.G. thanks the National Science Foundation for Graduate Fellowship support. P.Y. is an Alfred P. Sloan Research Fellow. Work at the Lawrence Berkeley National Laboratory was supported by the Office of Science, Basic Energy Sciences, Division of Materials Science of the US Department of Energy.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.Y.
(e-mail: p_yang@uclink.berkeley.edu).

Deep roots of the Messinian salinity crisis

Svend Duggen^{*†}, Kaj Hoernle^{*}, Paul van den Bogaard^{*}, Lars Rüpke^{*} & Jason Phipps Morgan^{*}

^{*}GEOMAR Research Center for Marine Geosciences, Wischhofstrasse 1-3, 24148 Kiel, Germany

The Messinian salinity crisis—the desiccation of the Mediterranean Sea between 5.96 and 5.33 million years (Myr) ago¹—was one of the most dramatic events on Earth during the Cenozoic era². It resulted from the closure of marine gateways between the Atlantic Ocean and the Mediterranean Sea, the causes of which remain enigmatic. Here we use the age and composition of volcanic rocks to reconstruct the geodynamic evolution of the westernmost Mediterranean from the Middle Miocene epoch to the Pleistocene epoch (about 12.1–0.65 Myr ago). Our data show that a marked shift in the geochemistry of mantle-derived volcanic rocks, reflecting a change from subduction-related to intraplate-type volcanism, occurred between 6.3 and 4.8 Myr ago, largely synchronous with the Messinian salinity crisis. Using a thermomechanical model, we show that westward roll back of subducted Tethys oceanic lithosphere and associated asthenospheric upwelling provides a plausible mechanism for producing the shift in magma chemistry and the necessary uplift (~1 km) along the African and Iberian continental margins to close the Miocene marine gateways, thereby causing the Messinian salinity crisis.

In the Late Miocene (~8 Myr ago), marine passages in southern Spain and northern Morocco linked the Mediterranean Sea to the Atlantic Ocean (Fig. 1). The palaeodepth of the Rifean corridor in Morocco—the deepest of these gateways—was estimated at 600–800 m (refs 3, 4). Marine sediments, such as terraced reef complexes from these former channels, now outcrop several hundred metres above sea level^{5,6}. Three possible mechanisms have been proposed to close these marine passages: (1) global sea level drop of ~60 m (refs 7, 8); (2) horizontal shortening associated with crustal nappe movements⁹; and (3) tectonic uplift^{3,4,10}.

Not only was the ~60-m drop in sea level in the Messinian insufficient to have closed all of the Late Miocene marine gateways, but also it has recently been shown that the onset of evaporite deposition at 5.96 ± 0.02 Myr does not correspond to the open-ocean benthic $\delta^{18}\text{O}$ signal, which is commonly interpreted to reflect glacio-eustatic sea-level changes. Therefore a global sea-level drop cannot have caused the Messinian salinity crisis (MSC)^{1,11}. Horizontal shortening is also an unlikely mechanism, for two reasons. First, because emplacement of crustal nappes in the Early Miocene, connected with extensional collapse of the Alborán block thickened through the collision of Africa and Iberia¹², had already ceased before the Late Miocene in the Betics¹³, and second, there is no evidence that nappe emplacement blocked the Rifean corridor in the Late Miocene (Fig. 1)³. It has, however, been shown that sediments in the former marine gateways were uplifted in the Late Miocene and Pliocene to their present elevations^{3,5,10}. Palaeodepth reconstructions indicate a rapid shallowing in the Late Miocene with rates as high as 5 mm yr⁻¹ at 7.2 Myr ago in the southern Rifean corridor, which was emergent by ~6.0 Myr (ref. 3). An increase of continental detritus in the uppermost Pliocene deposits of the Alborán basin indicates a continuation of this uplift into the Pliocene¹⁴.

[†] Present address: Department of Geology, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK.

The Betic and Rifan corridors are traversed by a north-north-east-trending volcanic belt, ~500 km long by ~200 km wide, extending from southeastern Spain through the eastern Alborán Sea into northwestern Africa (Fig. 1b). As volcanism is the surface expression of mantle processes, we use the age and geochemistry of volcanic rocks from this belt to trace the geodynamic evolution of the Alborán mantle, in order to elucidate the origin of uplift from the Late Miocene to Pliocene. We have dated more than 70 volcanic rocks, covering the entire Alborán volcanic province, using the laser ^{40}Ar – ^{39}Ar technique on amphibole, biotite and feldspar phenocrysts, microcrystalline rock matrix and glass separates. The dated samples were also analysed for major and trace elements, and Sr, Nd, Pb and O isotopic composition.

Two distinct groups of volcanic rocks occur throughout the eastern Alborán volcanic belt, as defined by their SiO_2 and alkali contents (Fig. 1b and 2a): (1) a Middle Miocene to Lower Pliocene (12.1–4.8 Myr ago) high-Si group, and (2) an Upper Miocene to Pleistocene (6.3–0.65 Myr ago) low-Si group. The older, high-Si, group can be further subdivided into two subgroups: a K-poor tholeiitic and calc-alkaline series of basalts through rhyolites with relatively low concentrations of incompatible elements, and K-rich shoshonites and lamproites with high concentrations of highly to moderately incompatible elements (Fig. 2c). The younger, low-Si volcanic rocks are alkalic, ranging from alkali basalt and basanite to hawaiite and tephrite.

Mafic members from both age groups also exhibit distinct trace-element and isotopic compositions. Whereas mafic rocks of the older high-Si group show relative enrichment in U, K and Pb and depletion in Nb and Ta (Fig. 2c), mafic rocks from the younger low-Si group show the opposite: relative depletion in U, K and Pb and enrichment in Nb and Ta (Fig. 2b). In addition, the low-Si group have lower Pb/Ce, K/La, K/Nb and $^{87}\text{Sr}/^{86}\text{Sr}$ but higher Nb/U,

$^{143}\text{Nd}/^{144}\text{Nd}$, $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ ratios (Fig. 3).

The major-element, trace-element and Sr–Nd–Pb–O isotope geochemistry of the older high-Si mafic volcanic rocks are characteristic of volcanism occurring in subduction zones (for example, the Izu and Aeolian arcs, Fig. 2c). Trace-element modelling and combined Sr and O isotope data provide the strongest evidence for subduction beneath the Alborán in the Middle and Late Miocene. The O and Sr isotopic compositions of plagioclase and clinopyroxene phenocrysts in the most mafic K-poor, high-Si rocks fall along a convex-down mixing curve between mid-ocean-ridge basalt (MORB) mantle and sediments on a $^{87}\text{Sr}/^{86}\text{Sr}$ versus $\delta^{18}\text{O}$ diagram, as do K-rich whole-rock samples from southern Spain¹⁵. These isotopic compositions thus cannot be explained through crustal assimilation¹⁶, which generates a convex-up mixing curve. The combined O and Sr isotopic data are therefore diagnostic of mantle source contamination with fluids or melts from subducted sediments^{17,18}. In addition, the K-rich subgroup and low-Si group commonly contain mantle xenoliths, also ruling out significant crustal assimilation in these rocks. Therefore the dramatic change in the composition of mafic, mantle-derived volcanic rocks in the eastern Alborán realm between 6.3 Myr ago (earliest alkali basalt) and 4.8 Myr ago (latest shoshonite) indicates that a major change in mantle dynamics occurred slightly before, or at the beginning of, this time interval.

The geochemistry of the younger low-Si mafic volcanic rocks are similar to Na-rich, alkalic intraplate basaltic rocks from the North Atlantic (for example, Canary Islands, Fig. 2b) and Africa (for example, Ahaggar)¹⁹ derived from asthenospheric sources. Intraplate-type volcanism results from the decompression melting of upwelling asthenospheric mantle. We however discount derivation of this volcanism from a mantle plume, because of the lack of an age progression in the younger low-Si volcanism, and because of

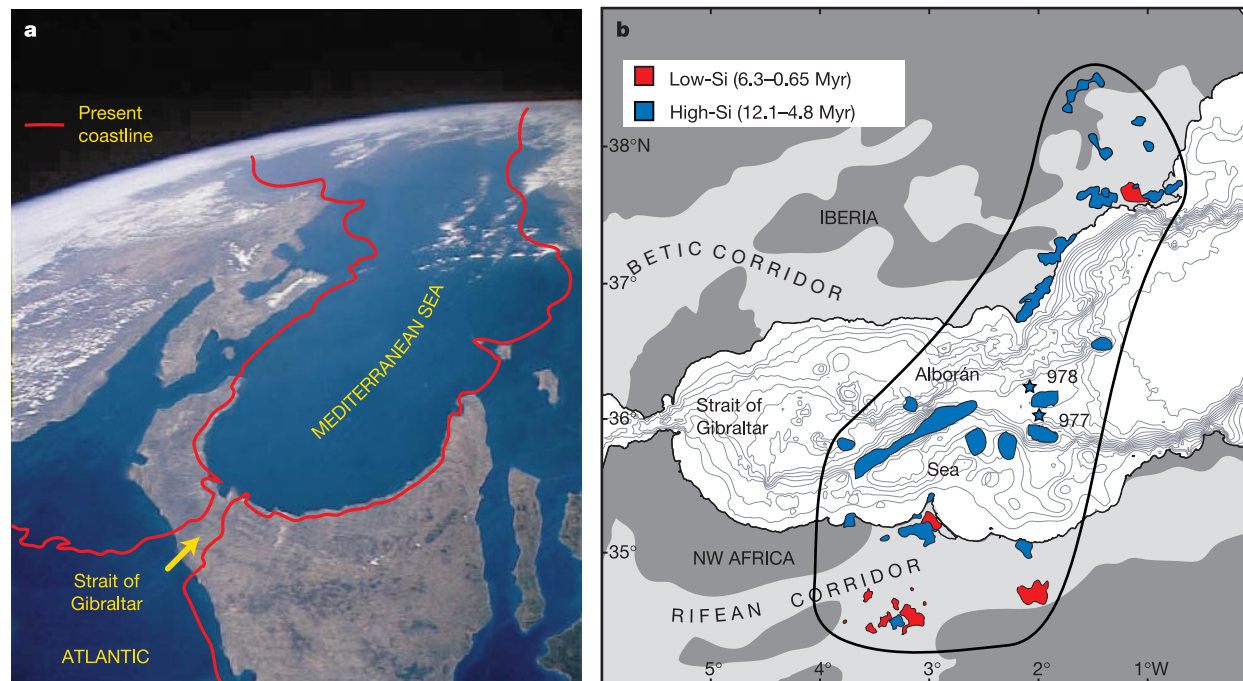


Figure 1 Maps illustrating Late Miocene Atlantic–Mediterranean marine gateways, and the distribution of Miocene–Pleistocene volcanism in the westernmost Mediterranean.

a, NASA photograph of the westernmost Mediterranean, modified to show Atlantic–Mediterranean marine gateways in southern Spain and northwestern Africa about 8 Myr ago, based on the distribution of Upper Miocene reef complexes and marine sediments⁵.

b, Map showing the westernmost Mediterranean and the 500 × 200 km north-northeast-

trending eastern Alborán volcanic belt crossing the marine gateways in both southern Spain (Betic corridor) and in northwestern Africa (Rifan corridor). Middle Miocene to Lower Pliocene high-Si group volcanic rocks are shown in blue, and Upper Miocene to Pleistocene low-Si group volcanic rocks are shown in red. Numbers 977 and 978 refer to drill holes during Ocean Drilling Program (ODP) Leg 161 (ref. 28).

lacking evidence in seismic tomographic studies for low-velocity cylindrical structures beneath southeastern Spain, the Alborán Sea and northern Morocco^{20,21}.

In fact, new seismic data provide compelling evidence for active east-dipping subduction beneath Gibraltar²⁰. Seismic tomography shows an ~250-km-wide, near-vertical slab extending from the surface to the boundary between the upper and lower mantle^{20,21}. In

a continental collisional setting, intraplate-type volcanism can also be generated by removal of a lithospheric body (oceanic lithosphere or subcontinental lithospheric mantle); for example, the Pannonian basin, eastern Europe²², and the Andes, South America²³. Westward roll back and steepening of the subducting slab in the Late Miocene would allow hotter, less-dense asthenosphere to ascend into the void left by the removal of the slab. Decompression melting of the upwelling asthenosphere can generate alkalic volcanism.

In order to determine if these mantle processes could produce the necessary uplift to close the Miocene marine gateways, we have done two-endmember, two-dimensional calculations to estimate the possible uplift resulting from roll back of the Tethys lithosphere (Fig. 4). With the modelling, we examine the effect of removing cold, dense material from the base of the Alborán lithosphere, replacing it with hot asthenosphere. For the current Gibraltar geometry, the 'slab-hinge effect' can lead to flexural-bending-induced uplift of ~100 m at the hinge, and uplift of ~300 m in the region where asthenosphere has infilled above the former slab (Fig. 4a). The removal of lithospheric mantle beneath the margins of

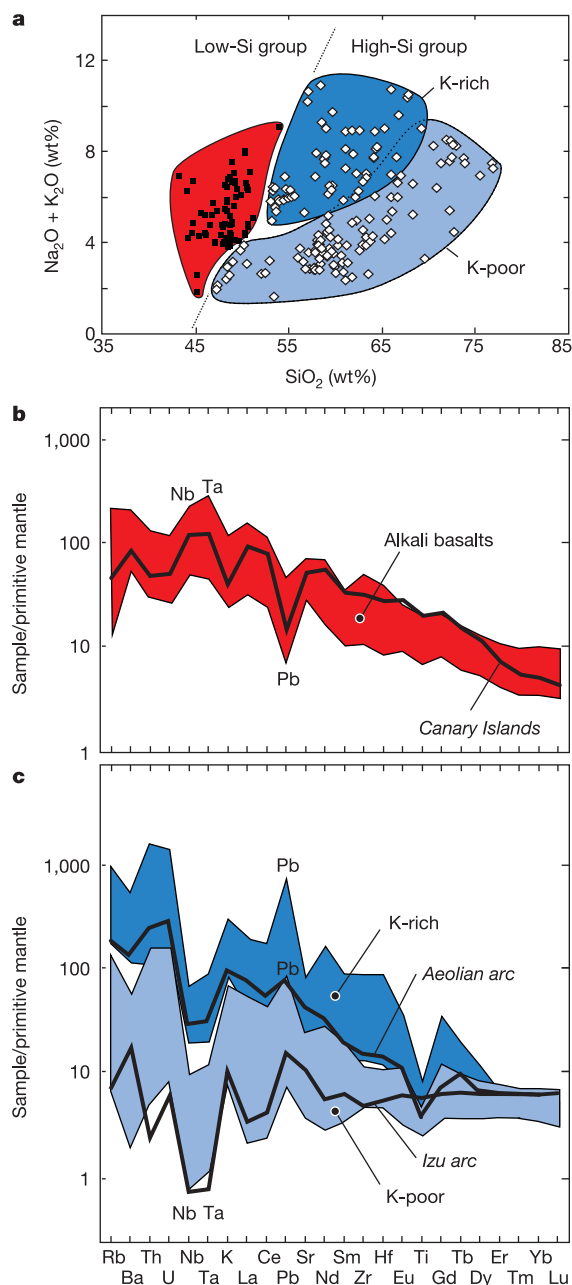


Figure 2 The major- and trace-element geochemistry of Miocene–Pleistocene volcanic rocks in the westernmost Mediterranean changes from subduction-related to intraplate-type during the Messinian. **a–c**, Major-element (**a**) and mantle-normalized trace-element diagrams (**b**, **c**) show the differences between the Upper Miocene to Pleistocene (6.3–0.65 Myr ago) low-Si group volcanic rocks (intraplate-type; **b**) and the most mafic Middle Miocene to Lower Pliocene (12.1–4.8 Myr ago) high-Si group volcanic rocks (subduction-related; **c**). Solid lines are representative samples from Canary Islands intraplate volcanic rocks in **b**, and Aeolian^{25,26} and Izu (Ta = Nb/17) arc²⁷ (subduction zone) magmas in **c**. Additional Alborán Sea data from ref. 28. Primitive mantle after ref. 29.

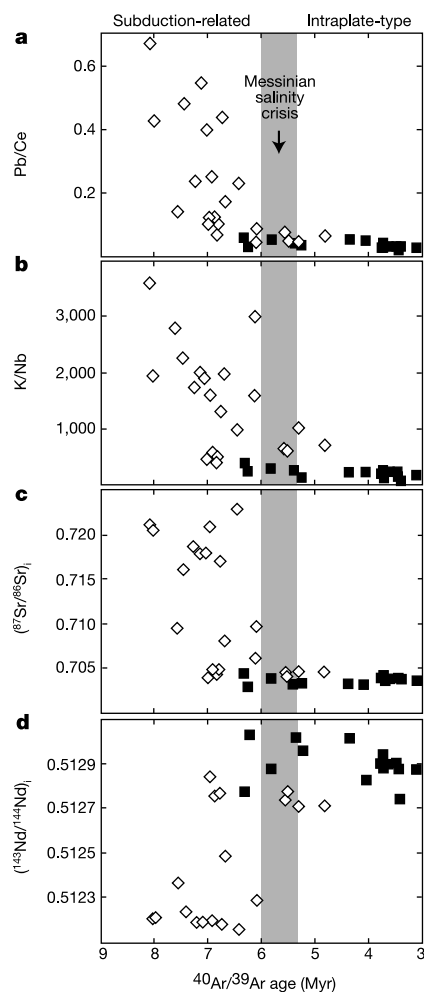


Figure 3 Trace-element and isotope ratios change during the MSC, consistent with the change from subduction-related to intraplate-type magmatism. Pb/Ce, K/Nb, initial ⁸⁷Sr/⁸⁶Sr and ¹⁴³Nd/¹⁴⁴Nd versus ⁴⁰Ar/³⁹Ar ages show a marked change between about 6.3 and 4.8 Myr ago, reflecting the shift from subduction-related volcanism (diamonds) to intraplate-type volcanism (squares). Intermediate compositions (<6.5 Myr) indicate interaction of intraplate-type magmas with the shallow mantle metasomatized by subduction processes, which was completed by ~4.8 Myr ago. Additional data sources, refs 16, 28, 30. Duration of MSC after ref. 1.

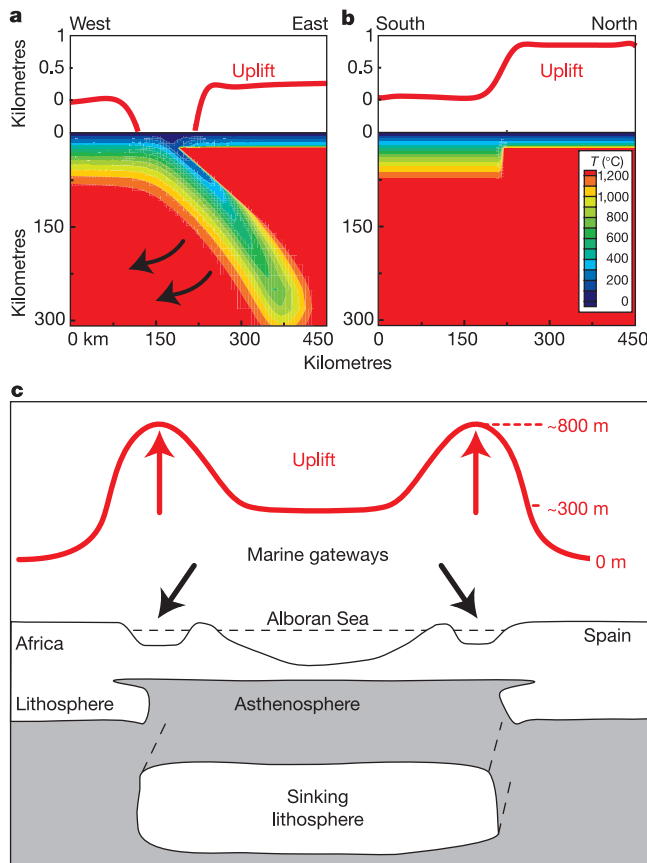


Figure 4 Thermomechanical models illustrating the uplift resulting from roll back of subducted Tethys oceanic lithosphere and associated asthenosphere upwelling as a plausible mechanism for the closure of Late Miocene Atlantic–Mediterranean marine gateways. **a**, ‘Slab-hinge effect’ from slab weight shown on a west–east profile; **b**, ‘edge-effect’ resulting from delamination of lithospheric mantle beneath the continental margins shown on a north–south profile. The edge-effect is larger in magnitude (~1 km). Relief is calculated assuming (1) a temperature-dependent viscous rheology, and (2) changes in the buoyancy resulting from replacement of cold oceanic lithosphere (blue and green) with hot 1,300 °C asthenosphere (red). The thermal model is based on the model formulation of ref. 31. To calculate the dynamic topography, we assume that the surface will deform with the uplifted ‘overburden’ being supported by flow-induced normal stresses³². **c**, A composite model for uplift in the Alborán region, combining the thermodynamical models of **a** and **b**. It illustrates that maximum uplift (~1 km) will occur on the continental margins, where the Late Miocene marine gateways linking the Mediterranean Sea to the Atlantic Ocean were located.

Morocco and southern Spain (the continental–ocean ‘edge-effect’) can lead to rapid uplift of the continental margins of the order of 800 m (Fig. 4b). We therefore propose that the roll back of Tethys oceanic lithosphere also peels away (delaminates) bands of lithospheric mantle from beneath the continental margins (Fig. 4c). Therefore, westward slab roll back is a plausible mechanism that can explain not only the location and composition of volcanism in the Alborán region, but also the uplift of the northern African and the southern Iberian margins, closing the Miocene gateways to the Atlantic Ocean.

The timing, location and magnitude of inferred mantle events and crustal uplift^{3,5,10} correlate well. Therefore the asthenospheric and crustal response (~6.3–4.8 Myr ago) to westward roll back of oceanic lithosphere (8–7 Myr ago) provides a possible trigger and driving force for the uplift that led to the exhumation of the marine gateways between the Atlantic and Mediterranean in the Late Miocene. The isolation of the Mediterranean Sea from the world-

wide network of oceans ultimately allowed for its desiccation during the MSC (5.96–5.33 Myr ago)¹. We note that the catastrophic flooding of the Mediterranean at the end of the MSC^{1,24} may also have had a mantle-related cause. Westward-migrating Late Miocene uplift may have also caused gravity-induced slumping from the western margin of the Gibraltar arc into the Atlantic abyssal plains, which—coupled with faulting—may have allowed a new marine gateway to open at the Strait of Gibraltar. □

Received 2 December 2002; accepted 27 February 2003; doi:10.1038/nature01553.

- Krijgsman, W., Hilgen, F. J., Raffi, I., Sierro, F. J. & Wilson, D. S. Chronology, causes and progression of the Messinian salinity crisis. *Nature* **400**, 652–655 (1999).
- Hsü, K. J., Ryan, W. B. F. & Cita, M. B. Late Miocene desiccation of the Mediterranean. *Nature* **242**, 240–244 (1973).
- Krijgsman, W. *et al.* Late Neogene evolution of the Taza–Guerfic Basin (Rifean Corridor, Morocco) and implications for the Messinian salinity crisis. *Mar. Geol.* **153**, 147–160 (1999).
- Hodell, D. A., Benson, R. H., Kennet, J. P. & El Bied, K. R. Stable isotope stratigraphy of Latest Miocene sequences in northwest Morocco: The Bou Regreg section. *Palaeogeography* **4**, 467–482 (1989).
- Esteban, M., Braga, J. C., Martin, J. & de Santisteban, C. in *Models for Carbonate Stratigraphy from Miocene Reef Complexes of Mediterranean Regions* (eds Franseen, E. K., Esteban, M., Ward, W. C. & Rouchy, J.-M.) 55–72 (SEPM, Society for Sedimentary Geology, Tulsa, 1996).
- Roger, S. *et al.* ⁴⁰Ar/³⁹Ar dating of the pre-evaporitic Messinian marine sequences of the Melilla basin (Morocco): a proposal for some biosedimentary events as isochrons around the Alboran Sea. *Earth Planet. Sci. Lett.* **179**, 101–113 (2000).
- Adams, C. G., Benson, R. H., Kidd, R. B., Ryan, W. B. F. & Wright, R. C. The Messinian salinity crisis and evidence of late Miocene eustatic changes in the world ocean. *Nature* **269**, 383–386 (1977).
- Hodell, D. A., Elmsstrom, K. M. & Kennet, J. P. Latest Miocene benthic $\delta^{18}\text{O}$ changes, global ice volume, sea level and the ‘Messinian salinity crisis’. *Nature* **320**, 411–414 (1986).
- Weijermars, R. Neogene tectonics in the Western Mediterranean may have caused the Messinian salinity crisis and an associated glacial event. *Tectonophysics* **148**, 211–219 (1988).
- Garcés, M., Krijgsman, W. & Agustí, J. Chronology of the late Turolian deposits of the Fortuna basin (SE Spain): implications for the Messinian evolution of the eastern Betics. *Earth Planet. Sci. Lett.* **163**, 69–81 (1998).
- Hodell, D. A., Curtis, J. H., Sierro, F. J. & Raymo, M. E. Correlation of late Miocene to early Pliocene sequences between the Mediterranean and North Atlantic. *Palaeogeography* **16**, 164–178 (2001).
- Platt, J. P. & Vissers, R. L. M. Extensional collapse of thickened continental lithosphere: A working hypothesis for the Alboran Sea and Gibraltar Arc. *Geology* **17**, 540–543 (1989).
- Comas, M. C., Platt, J. P., Soto, J. I. & Watts, A. B. The origin and tectonic history of the Alboran Basin. *Proc. ODP Sci. Res.* **161**, 555–580 (1999).
- Martínez-Ruiz, F., Comas, M. C. & Alonso, B. Mineral associations and geochemical indicators in Upper Miocene to Pleistocene sediments in the Alboran Basin. *Proc. ODP Sci. Res.* **161**, 21–36 (1999).
- Benito, R. *et al.* Sr and O isotope constraints on source and crustal contamination in the high-K calc-alkaline and shoshonitic neogene volcanic rocks of SE Spain. *Lithos* **46**, 733–802 (1999).
- Turner, S. P. *et al.* Magmatism associated with orogenic collapse of the Betic–Alboran Domain, SE Spain. *J. Petrol.* **40**, 1011–1036 (1999).
- James, D. E. The combined use of oxygen and radiogenic isotopes as indicators of crustal contamination. *Annu. Rev. Earth Planet. Sci.* **9**, 311–344 (1981).
- Taylor, H. P. Jr & Sheppard, S. M. F. in *Stable Isotopes in High Temperature Geologic Processes* (eds Valley, J. W., Taylor, H. P. Jr & O’Neil, J. R.) 227–271 (Reviews in Mineralogy 16, Mineralogical Society of America, Washington DC, 1986).
- Allège, C.-J., Dupré, B., Lambret, B. & Richard, P. The subcontinental versus suboceanic debate; I. Lead-neodymium-strontium isotopes in primary alkali basalts from a shield area; the Ahaggar volcanic suite. *Earth Planet. Sci. Lett.* **52**, 85–92 (1981).
- Gutscher, M.-A. *et al.* Evidence for active subduction beneath Gibraltar. *Geology* **30**, 1071–1074 (2002).
- Calvert, A. *et al.* Geodynamic evolution of the lithosphere and upper mantle beneath the Alboran region of the western Mediterranean: Constraints from travel time tomography. *J. Geophys. Res.* **105**, 10871–10898 (2000).
- Girbacea, R. & Frisch, W. Slab in the wrong place: Lower lithospheric mantle delamination in the last stage of the Eastern Carpathian subduction retreat. *Geology* **26**, 611–614 (1998).
- Kay, R. W. & Mahlburg Kay, S. Delamination and delamination magmatism. *Tectonophysics* **219**, 177–189 (1993).
- Hsü, K. J. *et al.* in *Initial Reports of the Deep Sea Drilling Project 42* (eds Hsü, K. J. *et al.*) 1053–1078 (US Government Printing Office, 1978).
- Ellam, R. M. *et al.* The transition from calc-alkaline to potassic orogenic magmatism in the Aeolian Islands, Southern Italy. *Bull. Volcanol.* **50**, 386–398 (1988).
- Ellam, R. M., Hawkesworth, C. J., Menzies, M. A. & Rogers, N. W. The volcanism of Southern Italy: Role of subduction and the relationship between potassic and sodic alkaline magmatism. *J. Geophys. Res.* **94**, 4589–4601 (1989).
- Taylor, R. N. & Nesbitt, R. W. Isotopic characteristics of subduction fluids in an intra-oceanic setting, Izu–Bonin Arc, Japan. *Earth Planet. Sci. Lett.* **164**, 79–98 (1998).
- Hoernle, K., Bogaard, P. v. d., Duggen, S., Mocek, B. & Garbe-Schönberg, D. Evidence for Miocene subduction beneath the Alboran Sea: ⁴⁰Ar/³⁹Ar dating and geochemistry of volcanic rocks from Holes 977A and 978A. *Proc. ODP Sci. Res.* **161**, 357–373 (1999).
- Hofmann, A. W. Chemical differentiation of the Earth. The relationship between mantle, continental crust, and oceanic crust. *Earth Planet. Sci. Lett.* **90**, 297–314 (1988).
- El Bakkali, S., Gourgaud, A., Bourdier, J.-L., Bellon, H. & Gundogdu, N. Post-collision neogene volcanism of the Eastern Rif (Morocco): magmatic evolution through time. *Lithos* **45**, 523–543 (1998).
- Rüpke, L. H., Phipps Morgan, J., Hort, M. & Conolly, J. A. D. Are the regional variations in Central American arc lavas due to differing basaltic versus peridotitic slab sources of fluids? *Geology* **30**, 1035–1038 (2002).
- Zong, S. & Gurnis, M. Controls on trench topography from dynamic models of subducted slabs. *J. Geophys. Res.* **99**, 15683–15695 (1994).

Acknowledgements We thank K. Reicherter, M. Hort and T. Hansteen for discussions, and M. Bouabdellah, M. Chaieb, D. Garbe-Schönberg, C. Harris, F. Hauff, M. Jadid, J. M. Fernandez Soler, M. Milhi, A. Moukadiri, D. Rau and J. Sticklus for analytical, field or logistic support. This work was supported by the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.D. (e-mail: s.duggen@gl.rhul.ac.uk).

N₂ production by the anammox reaction in the anoxic water column of Golfo Dulce, Costa Rica

Tage Dalsgaard*, Donald E. Canfield†, Jan Petersen†, Bo Thamdrup† & Jenaro Acuña-González‡

* Department of Marine Ecology, National Environmental Research Institute, Vejlsovej 25 PO Box 314, DK-8600 Silkeborg, Denmark

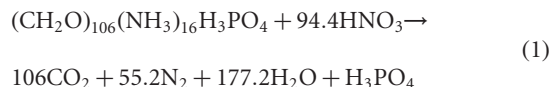
† Institute of Biology, Danish Centre for Earth System Science, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

‡ Centro de Investigación en Ciencias del Mar y Limnología, CIMAR, Universidad de Costa Rica, San José, Costa Rica

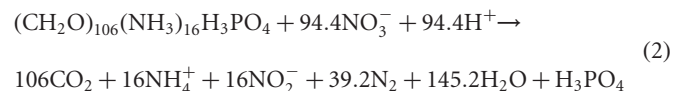
In oxygen-depleted zones of the open ocean, and in anoxic basins and fjords, denitrification (the bacterial reduction of nitrate to give N₂) is recognized as the only significant process converting fixed nitrogen to gaseous N₂. Primary production in the oceans is often limited by the availability of fixed nitrogen such as ammonium or nitrate¹, and nitrogen-removal processes consequently affect both ecosystem function and global biogeochemical cycles. It was recently discovered that the anaerobic oxidation of ammonium with nitrite—the ‘anammox’ reaction, performed by bacteria—was responsible for a significant fraction of N₂ production in some marine sediments². Here we show that this reaction is also important in the anoxic waters of Golfo Dulce, a 200-m-deep coastal bay in Costa Rica, where it accounts for 19–35% of the total N₂ formation in the water column. The water-column chemistry in Golfo Dulce is very similar to that in oxygen-depleted zones of the oceans—in which one-half to one-third of the global nitrogen removal is believed to occur^{3,4}. We therefore expect the anammox reaction to be a globally significant sink for oceanic nitrogen.

Golfo Dulce is connected to the equatorial tropical North Pacific Ocean through a 14-km-wide opening with a sill at 60 m depth. The bay is about 50 km long, with anoxic, nitrate-rich water in the bottom depths of the basin^{5,6}. During the sampling reported here, the pycnocline was located at 40–55 m depth, and the water column was anoxic at depths greater than about 100 m, although sulphide did not accumulate until the bottom 10–20 m (Fig. 1a). Nitrate reduction, and not sulphate reduction, dominated water-column processes through most of the anoxic zone, similar to well-developed oxygen-minimum zones in the eastern Pacific and Indian oceans^{7–9}. Curiously, ammonium (a normal product of anoxic organic matter mineralization) did not accumulate in the anoxic waters until nitrate was depleted and H₂S appeared (Fig. 1a). This observation alone indicates a significant role for anaerobic ammonium oxidation, where the ammonium liberated in the water column during denitrification is further oxidized with nitrate

(through nitrite) as shown in equation (1) (refs 10, 11):



This stoichiometry can be explained by the simultaneous operation of two known bacterial processes, denitrification (equation (2)), and the anammox reaction (equation (3)) (refs 2, 12, 13).



The stoichiometric expression for denitrification (equation (2)) is somewhat modified from the ‘standard’ formulation^{10,11}, showing here the production of nitrite, a free intermediate of denitrification¹⁴, in equal proportions to ammonium liberation. This is because anammox bacteria utilize nitrite and ammonium in a 1:1 ratio to form N₂ gas (equation (3)). It follows from equations (2) and (3) that the anammox reaction is responsible for 29% of the total N₂ production if all of the ammonium liberated during denitrification is subsequently oxidized by anammox. This possibility was recognized many years ago^{10,11}, although it was not then known whether organisms in nature could in fact conduct the anammox process.

We used ¹⁵N-labelling techniques to explore for anammox in the water column of Golfo Dulce. Samples were retrieved from four depths in the anoxic water column at two sites (stations A and B) in the 200-m-deep central part of the basin, and amended with different combinations of ¹⁵N- and ¹⁴N-labelled nitrate and ammonium. At both sites and all depths, we observed anaerobic oxidation of NH₄⁺ to N₂, as illustrated by the formation of ¹⁴N¹⁵N from added ¹⁵NH₄⁺ (Fig. 2a). The formation of ¹⁴N¹⁵N started immediately after addition of ¹⁵NH₄⁺, and the rate was constant over time. The formation of ¹⁵N¹⁵N was at the limit of detection. Therefore, the added ¹⁵NH₄⁺ was combined with nitrogen from the natural ¹⁴NO₃⁻ or ¹⁴NO₂⁻ pools to form N₂, which is characteristic of the anammox reaction^{12,13}. It has been suggested that anaerobic

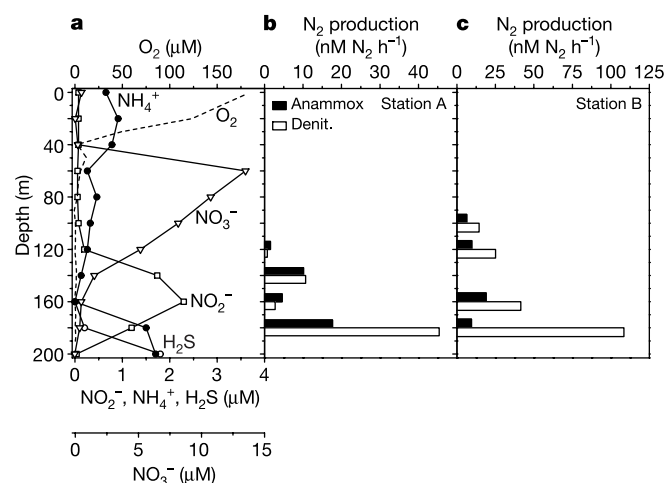


Figure 1 Water column data. **a**, Concentrations of O₂ (dashed line), NO₃⁻ (open triangles), NO₂⁻ (open squares), NH₄⁺ (filled circles) and H₂S (open circles) versus depth. The data presented are from Station B, but the water chemistry at Station A was almost identical. The main differences are that at Station A the NO₂⁻ peak only reached 0.6 μM, and H₂S was only detected at 200 m depth. **b**, **c**, Rates of anammox (black bars) and denitrification (white bars) at Station A (**b**) and Station B (**c**). Note the different scales in **b** and **c**. Rates are determined as averages of three incubations.

ammonium oxidation could occur with manganese oxide as oxidant producing either N_2 or nitrate^{15–17}. Under denitrifying conditions the nitrate would subsequently be reduced to N_2 . However, those mechanisms cannot explain the oxidation of ammonium to N_2 in this study, as both mechanisms would lead to the formation of $^{15}N^{15}N$ —which was not observed. Furthermore, if ammonium was oxidized to nitrate with manganese oxide, this would lead to a gradual increase in the ^{15}N abundance in the nitrate pool and the rate of production of $^{14}N^{15}N$ would thus increase over time. The linearity of the $^{14}N^{15}N$ production from $^{15}NH_4^+$ is incompatible with that mechanism (Fig. 2a).

With the addition of $^{15}NO_3^-$, both $^{14}N^{15}N$ and $^{15}N^{15}N$ were produced with rates increasing during the incubations (Fig. 2b, c). This time dependence is explained by the gradual build-up of ^{15}N in the nitrite pool. Initial concentrations of unlabelled nitrite were 0.5–1.3 μM in all incubations, except for Station B at 160 m where it was 2.4 μM . Nitrite is the electron acceptor for anammox, and a free intermediate in N_2 formation by denitrification, and added $^{15}NO_3^-$ must be transferred into the nitrite pool before ^{15}N -labelled N_2 can be produced.

In incubations with $^{15}NO_3^-$ and unlabelled, native NH_4^+ and $NO_3^- + NO_2^-$, anammox will produce $^{14}N^{14}N$ and $^{14}N^{15}N$, whereas denitrification produces $^{14}N^{14}N$, $^{14}N^{15}N$ and $^{15}N^{15}N$ through random isotope pairing^{2,12,13,18}. In our experiments, the production rate of $^{14}N^{15}N$ was considerably higher than expected for denitrification alone, indicating another source for $^{14}N^{15}N$, consistent with anammox activity. For the experiment depicted in Fig. 2b, a contribution of anammox to total N_2 production of 28% was determined from the specific labelling of ^{15}N in the nitrate pool, and the production rates of $^{14}N^{15}N$ and $^{15}N^{15}N$ (ref. 2). As the propagation of the ^{15}N label from nitrate to nitrite was greatest at the end of the experiment, the production rates of $^{14}N^{15}N$ and $^{15}N^{15}N$ were calculated as the slope of the concentration curves through the data points from the last two samplings. Rates of both anammox and denitrification may be slightly overestimated owing to the elevated nitrate concentrations in the $^{15}NO_3^-$ amendment

experiments. However, owing to the tight coupling of the two processes, and the NH_4^+ limitation of the anammox process, the relative importance of the two processes should not be affected.

At the three upper sampling depths of both stations, the ammonium concentration was around the detection limit (0–0.3 μM , Fig. 1a), and the low ammonium concentration apparently limited the rates of anammox. The limitation of ammonium on rates of anammox was substantiated by a stimulation of anammox when unlabelled ammonium was added together with the $^{15}NO_3^-$ (Fig. 2c). At the deepest sampling depth, where ammonium concentrations were higher, ammonium limitation was less pronounced, as indicated by either no stimulation of ammonium oxidation (Station B) or a much lower stimulation than in the three upper depths (Station A). Integrated through the whole anoxic anoxic water column, the ammonium addition stimulated anammox by a factor of 4 at Station A and by a factor of 2 at Station B. As anammox is ammonium limited in Golfo Dulce, there is the capacity for anammox organisms to oxidize all the ammonium produced by carbon mineralization in the anoxic water column. Therefore, the ammonium deficiency in the anoxic water column can be explained by the coupling between denitrification and anammox as expressed in equations (2) and (3).

Rates of anammox and denitrification were calculated from the $^{15}NO_3^-$ labelling experiments (Fig. 1b, c), as anammox in the $^{15}NH_4^+$ experiments was stimulated by the added ammonium (see above). The relative importance of anammox was greatest in the upper three sampling depths, where it accounted for on average 58% and 32% of the total N_2 production on stations A and B, respectively. The significance of anammox thus matched, or even exceeded, what is predicted by the equations above. Higher than expected proportions of anammox may be explained by the preferential decomposition of nitrogen-rich organic matter¹⁹ through denitrification, releasing more ammonia than predicted by standard Redfield stoichiometry (equation (2)). At the deepest sampling depth, anammox accounted for only 28% and 13% of total N_2 production at stations A and B, respectively. This relatively low contribution of anammox, particularly at Station B, could be a result of significant denitrification coupled to sulphide oxidation (Fig. 1a). Alternatively sulphide could inhibit the anammox process, but this has not been tested. Integrating the rates for both anammox and denitrification for the whole anoxic water column, we find that anammox contributed between 19% (Station B) and 35% (Station A) of the N_2 production (Table 1).

Our experiments demonstrate a tight coupling between the ammonium liberated during denitrification, and further transformation of this ammonium by the anammox process. Such a coupling has been considered¹⁰, but has not previously been documented experimentally in a marine water column. The water chemistry of Golfo Dulce has many similarities to oxygen-depleted oxygen-minimum zones, which also display extensive nitrate reduction, but without the build-up of ammonium^{4,17,20–22}. We therefore suspect that anammox is coupled to denitrification, and is responsible for ammonium oxidation in those environments as well. By current estimates, between about one-third and one-half of the total denitrification in the oceans occurs in oxygen-minimum zones^{3,4}.

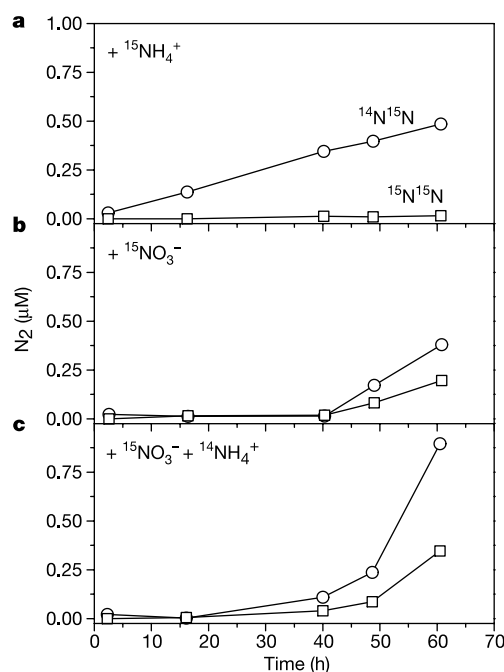


Figure 2 Isotopic enrichment experiments. **a–c**, Concentrations of $^{14}N^{15}N$ (open circles) and $^{15}N^{15}N$ (squares) in experiments with addition of $^{15}NH_4^+$ (**a**), $^{15}NO_3^-$ (**b**) and $^{15}NO_3^- + ^{14}NH_4^+$ (**c**). The concentrations of NO_3^- and NH_4^+ were 7.2 μM and 0.2 μM , respectively, before amendments. Data from Station B at 120 m depth. Each data point represents the average of three incubations.

Table 1 Integrated N_2 production rates

Station	Anammox (mmol $N m^{-2} d^{-1}$)	Denitrification (mmol $N m^{-2} d^{-1}$)	Relative anammox
A	42	79	35%
B	61	266	19%

Rates were calculated from experiments with $^{15}NO_3^-$ addition and integrated for the whole anoxic water column. Relative anammox is calculated as anammox/(anammox + denitrification). Depth-integrated rates were calculated as: $IR = \sum_{n=1}^{14} (R_n h_n)$ where IR is the integrated rate, R_n is the measured rate for depth interval n , h_n is the height of the water column for interval n . The boundaries for each interval was set as the midpoint between two consecutive measurements, the upper boundary was set at 100 m (Station A) or 90 m (Station B), and the lower boundary was set at 200 m.

Therefore, anammox in these waters could account for 10–15% of the marine N_2 production.

The anammox process may be more important than this, with particularly favourable conditions occurring in upwelling areas where anoxic nitrate-rich water reaches the sediment. A good example would be upwelling areas off the Peruvian and Chilean coasts, where sedimentary sulphide is oxidized by sulphide-oxidizing bacteria with nitrate, producing ammonium²³. This ammonium is released to the water column, probably enhancing the significance of anammox, and N_2 production, in these areas. Thus, the anammox process, with its distinct regulatory characteristics, should be included in studies of nitrogen cycling in the marine environment. □

Methods

Sampling was performed at Station A (8° 34.15' N, 83° 14.69' W) and Station B (8° 37.99' N, 83° 20.62' W) of Golfo Dulce in November 2001. Profiles of salinity, temperature and oxygen were measured at 10-m intervals with a DataSonde 4 (Hydrolab). All water samples were retrieved with a 5-l Niskin bottle (KC Denmark). For nutrient profiles, water was sampled from the Niskin bottle with a plastic syringe and filtered through a cellulose acetate filter (pore size 0.22 µm) into polypropylene vials that were stored on ice until return to the laboratory where they were frozen for later analysis. For the ^{15}N -labelling experiments, water was sampled at 120, 140, 160 and 180 m depth at Station A, and at 100, 120, 160 and 180 m depth at Station B. Water was transferred from the Niskin bottle via Tygon tubing into the bottom of a 250-ml glass bottle and allowed to flow over for half a volume change. The bottle was closed with a Viton stopper taking care to exclude bubbles, and stored on ice until return to the laboratory. Experiments were started no later than 6 h after sampling.

For each of the four depths from each station experiments were started by the addition of ^{15}N labelled and unlabelled nitrate and ammonium to the 250-ml bottles to the following final concentrations from concentrated stock solutions: $10 \mu M$ $^{15}NO_3^-$, $10 \mu M$ $^{15}NO_3^- + 10 \mu M$ $^{14}NH_4^+$ and $10 \mu M$ $^{15}NH_4^+$. After addition the amended water was bubbled with helium for 15 min to facilitate the detection of N_2 production, and then transferred to 12-ml glass vials with a 5-mm butyl rubber septum (Exetainers, Labco) through a glass tube leading from the bottle into the bottom of the Exetainer and allowed to overflow. The Exetainers were incubated at *in situ* temperature in a water bath kept anoxic with an alkaline sodium ascorbate solution. The oxygen concentration in the Exetainers was measured with a microelectrode designed for measuring low oxygen concentrations (Unisense) and was below the detection limit of 0.2 µM throughout the experiment. At each time point Exetainers were sampled in triplicate by removing 5 ml of water while replacing it with helium and adding 0.1 ml of 50% (w/v) ZnCl to stop biological activity in the Exetainer. The removed water was filtered for nutrient analysis and stored as described above.

The concentration of $NO_3^- + NO_2^-$ was determined using the vanadium chloride reduction method²⁴ (NO_x analyser model 42C, Thermo Environmental Instruments Inc.). Nitrite was analysed spectrophotometrically²⁵ and NH_4^+ was determined using the flow injection method with conductivity detection²⁶. The isotopic composition of the nitrate and ammonium pool was estimated from their concentrations before and after amendment. Concentrations of $^{14}N^{15}N$ and $^{15}N^{15}N$ were determined by isotope ratio mass spectrometry and calculated as excess above their natural abundance².

Received 28 October 2002; accepted 3 March 2003; doi:10.1038/nature01526.

- Falkowski, P. G., Barber, R. T. & Smetacek, V. Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**, 200–206 (1998).
- Thamdrup, B. & Dalsgaard, T. Production of N_2 through anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments. *Appl. Environ. Microbiol.* **68**, 1312–1318 (2002).
- Gruber, N. & Sarmiento, J. L. Global patterns of marine nitrogen fixation and denitrification. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
- Codispoti, L. A. *et al.* The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene? *Sci. Mar.* **65** (suppl. 2), 85–105 (2001).
- Thamdrup, B., Canfield, D. E., Ferdelman, T. G., Glud, R. N. & Gundersen, J. K. A biogeochemical survey of the anoxic basin Golfo Dulce, Costa Rica. *Rev. Biol. Trop.* **44**, 19–33 (1996).
- Vargas, J. A. Pacific coastal ecosystems of Costa Rica with emphasis on the Golfo Dulce and adjacent areas: A synoptic view based on the R.V. Victor Hensen expedition 1993/94 and previous studies. Preface. *Rev. Biol. Trop.* **44**, U1–U4 (1996).
- Lipschultz, F. *et al.* Bacterial transformations of inorganic nitrogen in the oxygen-deficient waters of the Eastern Tropical South-Pacific ocean. *Deep-Sea Res.* **37**, 1513–1541 (1990).
- Bange, H. W. *et al.* A revised nitrogen budget for the Arabian Sea. *Glob. Biogeochem. Cycles* **14**, 1283–1297 (2000).
- Deutsch, C., Gruber, N., Key, R. M., Sarmiento, J. L. & Ganachaud, A. Denitrification and N_2 fixation in the Pacific Ocean. *Glob. Biogeochem. Cycles* **15**, 483–506 (2001).
- Richards, F. A. in *Advances in Water Pollution Research* (ed. Pearson, E. A.) 215–232 (Pergamon, London, 1965).
- Richards, F. A. in *Chemical Oceanography* (eds Riley, J. P. & Skirrow, G.) 611–645 (Academic, London, 1965).
- van de Graaf, A. *et al.* Anaerobic oxidation of ammonium is a biologically mediated process. *Appl. Environ. Microbiol.* **61**, 1246–1251 (1995).
- Dalsgaard, T. & Thamdrup, B. Factors controlling anaerobic ammonium oxidation with nitrite in marine sediments. *Appl. Environ. Microbiol.* **68**, 3802–3808 (2002).
- Knowles, R. Denitrification. *Microbiol. Rev.* **46**, 43–70 (1982).

- Luther, G. W. I., Sundby, B., Lewis, P. J. & Silverburg, N. Interactions of manganese with the nitrogen cycle: alternative pathways to dinitrogen. *Geochim. Cosmochim. Acta.* **61**, 4043–4052 (1997).
- Hulth, S., Aller, R. C. & Gilbert, F. Coupled anoxic nitrification/manganese reduction in marine sediments. *Geochim. Cosmochim. Acta.* **63**, 49–66 (1999).
- Murray, J. W., Lee, B., Bullister, J. & Luther, G. W. in *Environmental Degradation of the Black Sea: Challenges and Remedies* (eds Besiktepe, S. T., Ünlüata, Ü. & Bologna, A. S.) 75–91 (Kluwer Academic, Dordrecht, 1999).
- Nielsen, L. P. Denitrification in sediment determined from nitrogen isotope pairing. *FEMS Microbiol. Ecol.* **86**, 357–362 (1992).
- Van Mooy, B. A. S., Keil, R. G. & Devol, A. H. Impact of suboxia on sinking particulate organic carbon: Enhanced carbon flux and preferential degradation of amino acids via denitrification. *Geochim. Cosmochim. Acta.* **66**, 457–465 (2002).
- Codispoti, L. A. & Christensen, J. P. Nitrification, denitrification and nitrous oxide cycling in the Eastern tropical South Pacific Ocean. *Mar. Chem.* **16**, 277–300 (1985).
- Cline, J. D. & Richards, F. A. Oxygen deficient conditions and nitrate reduction in the Eastern tropical northern Pacific Ocean. *Limnol. Oceanogr.* **17**, 885–900 (1972).
- Naqvi, S. W. A. Some aspects of the oxygen-deficient conditions and denitrification in the Arabian Sea. *J. Mar. Res.* **45**, 1049–1072 (1987).
- Otte, S. *et al.* Nitrogen, carbon, and sulfur metabolism in natural Thioploca samples. *Appl. Environ. Microbiol.* **65**, 3148–3157 (1999).
- Braman, R. S. & Hendrix, S. A. Nanogram nitrite and nitrate determination in environmental and biological materials by vanadium (III) reduction with chemiluminescence detection. *Anal. Chem.* **61**, 2715–2718 (1989).
- Grasshoff, K., Ehrhardt, M. & Kremling, K. *Methods of Seawater Analysis* (Verlag Chemie, Weinheim, 1983).
- Hall, P. O. J. & Aller, R. C. Rapid, small-volume, flow injection analysis for CO_2 and NH_4^+ in marine and freshwaters. *Limnol. Oceanogr.* **37**, 1113–1119 (1992).

Acknowledgements We thank J. A. Vargas for assistance in arranging field work, and E. Ruiz and D. Morera for assistance during sampling. We also thank L. Salling, P. Søholt, K. G. Lauridsen, E. Frandsen, T. Quottrup, M. V. Skjærbæk and A. Haxen for analytical work. D.E.C., J.P. and B.T. were supported by the Danish National Research Foundation and the Danish National Science Research Council.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.D. (e-mail: tda@dmu.dk).

Anaerobic ammonium oxidation by anammox bacteria in the Black Sea

Marcel M. M. Kuypers*, Olav Sliekers†, Gaute Lavik*, Markus Schmid†, Bo Barker Jørgensen*, J. Gijs Kuenen†, Jaap S. Sinninghe Damsté‡, Marc Strous§ & Mike S. M. Jetten§

* Max Planck Institute for Marine Microbiology (MPI), Department of Biogeochemistry, Celsiusstrasse 1, 28359 Bremen, Germany

† Department of Microbiology, Delft University of Technology, Julianalaan 67, 2628 BC Delft, The Netherlands

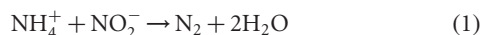
‡ Royal Netherlands Institute for Sea Research (NIOZ), Department of Marine Biogeochemistry and Toxicology, PO Box 59, 1790 AB Den Burg, The Netherlands

§ Department of Microbiology, University of Nijmegen, Toernooiveld 1, 6526 ED Nijmegen, The Netherlands

The availability of fixed inorganic nitrogen (nitrate, nitrite and ammonium) limits primary productivity in many oceanic regions¹. The conversion of nitrate to N_2 by heterotrophic bacteria (denitrification) is believed to be the only important sink for fixed inorganic nitrogen in the ocean². Here we provide evidence for bacteria that anaerobically oxidize ammonium with nitrite to N_2 in the world's largest anoxic basin, the Black Sea. Phylogenetic analysis of 16S ribosomal RNA gene sequences shows that these bacteria are related to members of the order Planctomycetales performing the anammox (anaerobic ammonium oxidation) process in ammonium-removing bio-reactors³. Nutrient profiles, fluorescently labelled RNA probes, ^{15}N tracer experiments and the distribution of specific 'ladder-ane' membrane lipids⁴ indicate that ammonium diffusing

upwards from the anoxic deep water is consumed by anammox bacteria below the oxic zone. This is the first time that anammox bacteria have been identified and directly linked to the removal of fixed inorganic nitrogen in the environment. The widespread occurrence of ammonium consumption in suboxic marine settings^{5–7} indicates that anammox might be important in the oceanic nitrogen cycle.

The Black Sea is the world's largest anoxic basin and is a model for both modern and ancient anoxic environments. It is characterized by a high ammonium concentration in the deep waters (up to 100 μM), whereas only trace amounts of fixed inorganic nitrogen are present in the 'suboxic' zone^{6,8} where the reduction of nitrate, manganese oxide or iron oxide occurs⁹. This apparent ammonium sink in the suboxic zone strongly suggests^{7,10,11} that ammonium is oxidized anaerobically to N_2 . Indeed, bacteria able to oxidize ammonia anaerobically have recently been discovered in laboratory bioreactors and wastewater treatment systems^{3,12}. These so-called 'anammox' bacteria belonging to the order Planctomycetales directly oxidize ammonia to N_2 with nitrite as the electron acceptor (Fig 1a, b):



During an R/V *Meteor* cruise in December 2001 we investigated the role of anammox in the Black Sea water column by using microbiological and biogeochemical techniques. In accord with earlier studies^{6,8,11} we observed a nitrate maximum at the bottom of the oxic zone in the western basin (site 7605; 42° 30.71' N, 30° 14.69' E; Fig. 2a). This maximum is caused by the mineralization of phytoplankton-derived organic nitrogen coupled to aerobic nitrification (Fig 1a). Ammonium concentrations are high in deep waters but decrease to background values above 97 m water depth (Fig. 2a). Aerobic nitrification cannot account for the consumption of ammonium because O_2 is absent below 80 m (Fig. 2b). However,

nitrate penetrates 15 m deeper in the water column, indicating that nitrate could be the oxidizer of ammonium¹¹. Alternatively, anammox bacteria could be using nitrite instead of nitrate to oxidize ammonium. Nitrite is an intermediate of denitrification and a nitrite peak is present at the base of the nitrate peak (Fig. 2a). Anammox in the suboxic zone could be coupled to nitrate reduction to nitrite (Fig 1a) by denitrifiers¹³, similarly to the process in anammox bioreactors¹⁴.

To check for anammox activity in the suboxic zone we anaerobically incubated water samples from various depths after the addition of [^{14}N]nitrite and [^{15}N]ammonium. Because the anammox process combines 1 mol of [^{15}N]ammonium and 1 mol of [^{14}N]nitrite to form 1 mol of single-labelled dinitrogen gas ($^{14}\text{N}^{15}\text{N}$) (equation (1)), the depth distribution of $\delta^{14}\text{N}^{15}\text{N}$ (Fig. 2c) expresses the potential anammox activity. The $\delta^{14}\text{N}^{15}\text{N}$ record shows a clear peak in the zone of nitrite and ammonium disappearance, whereas no significant anammox activity is observed outside the suboxic zone.

Specific biomarkers, so-called ladderane lipids, were used to trace anammox bacteria in particulate organic matter collected from various depths across the suboxic zone. Ladderane lipids⁴ are the main building blocks of a unique bacterial membrane that surrounds the anammoxosome, a special compartment of the anammox cell, in which the anaerobic oxidation of ammonium to N_2 takes place (Fig. 1b). Three different ladderane lipids were detected in the saponified total lipid extracts with a depth distribution (Fig. 2d and e) similar to that of the potential anammox activity (Fig. 2c), indicating that anammox bacteria could indeed be responsible for the anaerobic oxidation of ammonium. A clone library was generated from DNA extracted from Black Sea water at the depth of maximum ladderane abundance (90 m), after amplification of the 16S ribosomal RNA gene with primers specific for Planctomycetes¹⁵. Phylogenetic analysis of the 16S rRNA sequences confirms that the Planctomycetes, tentatively named *Candidatus* 'Scalindua sorokinii', from the suboxic zone of the Black Sea are related to bacteria known to be capable of the anammox process (87.9% sequence similarity to *Kuenenia*, 87.6% to *Brocadia*; Fig. 3). In fact, the sequence obtained from the Black Sea is nearly identical (98.1%) to a sequence recently obtained from a bioreactor shown to have anammox activity (M. Schmid, K. Walsh, R. Webb, W. I. Rijpstra, K. T. van de Pas Schoonen, T. C. J. Hill, B. F. Moffett, J. A. Fuerst, J.S.S.D., J. A. Harris, P. J. Shaw, M.S.M.J. and M. Strauss, unpublished observations). On the basis of the sequence obtained from the Black Sea, we designed an oligonucleotide probe, labelled with Cy3 fluorochrome, for fluorescence *in situ* hybridization (FISH). This probe gave a bright and specific signal with cells that have the unusual doughnut shape characteristic for anammox bacteria in bioreactors. Ladderane biomarkers and cells hybridizing with the new FISH probe (Fig. 1c) were also found in the suboxic zone at the shelf break (Station 7617, 43° 38.04' N, 30° 02.54' E), indicating that anammox bacteria are not restricted to the strongly stratified central basin but are also present in the more dynamic peripheral current⁸. The combined results clearly indicate that anammox bacteria are abundant and active in the Black Sea. Could these anammox bacteria be responsible for the observed ammonium sink in the suboxic zone of the Black Sea?

If we assume that the concentration profile of ammonium represents a steady state, an anaerobic ammonium oxidation rate of $\sim 0.007 \mu\text{M day}^{-1}$ was calculated for the suboxic zone of the central basin by using a reaction diffusion model. This rate is comparable to aerobic ammonium oxidation rates ($0.005\text{--}0.05 \mu\text{M day}^{-1}$) determined for the nitrate maximum of the western central basin of the Black Sea¹⁶. An anammox rate of 2–20 fmol ammonium per cell per day was found in laboratory bioreactors³. Assuming a similar range of cell-specific activity for the Black Sea, 300–3,000 anammox cells ml^{-1} would be needed to account for the observed ammonium oxidation rates in the suboxic zone. Counts of

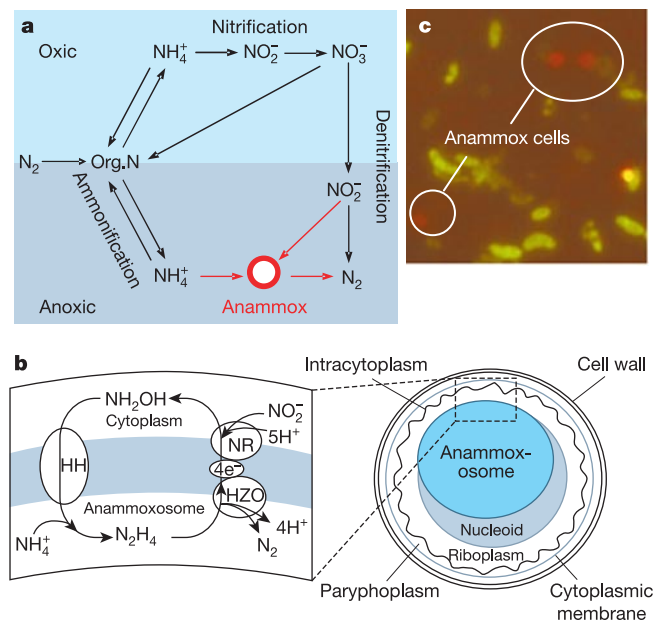


Figure 1 Morphology and physiology of anammox bacteria and their role in the marine nitrogen cycle. **a**, Simplified marine nitrogen cycle including the anammox 'sink'. Org.N, organic nitrogen. **b**, Morphology of the anammox cell and proposed model for the anammox process. HH, hydrazine (N_2H_4) hydrolase; HZO, hydrazine oxidizing enzyme; NR, nitrite reducing enzyme. **c**, Fluorescence *in situ* hybridization of filter material from station 7617 (142 m water depth). Green cells are total Eubacteria stained with EUB338 probe; red cells (encircled) are anammox bacteria stained with a new specific probe (AmxBS820).

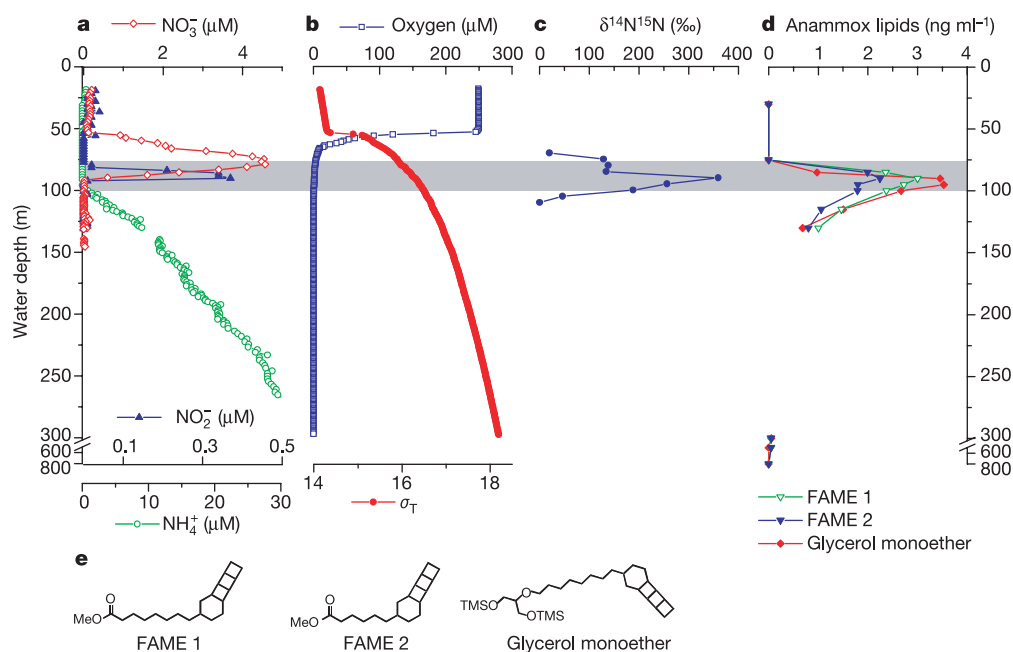


Figure 2 Chemical zoning and distribution of anammox indicators across the Black Sea chemocline. **a**, Fixed inorganic nitrogen species; **b**, water density and oxygen concentrations; **c**, interface peak of potential anammox activity expressed as anaerobic $^{15}\text{NH}_4^+$ oxidation by $^{14}\text{NO}_2^-$ to $^{14}\text{N}^{15}\text{N}$; **d**, peak of three ladderane membrane lipids specific for anammox bacteria; **e**, molecular structures of the three ladderane membrane lipids specific for anammox bacteria presented in **d**. The suboxic zone is indicated by grey

shading. Density (σ_T , the density of seawater in $\text{kg m}^{-3} - 1,000$), nitrate (NO_3^-), nitrite (NO_2^-), ammonium (NH_4^+) and oxygen profiles from Station 7605 ($42^\circ 30.71' \text{ N}$, $30^\circ 14.69' \text{ E}$). Ladderane lipid data from Stations 7605 and 7620 ($42^\circ 55.56' \text{ N}$, $30^\circ 03.65' \text{ E}$) were used to create a composite plot for the ladderane glycerol monoether and for the fatty acid methyl esters (FAMES) 1 and 2.

cells stained with the newly designed FISH probe (Amxbs820) gave an anammox cell density of $\sim 1,900 \pm 800 \text{ cells ml}^{-1}$ (0.75% of all cells counted by 4,6-diamidino-2-phenylindole (DAPI)) at the nitrite peak.

Although we acknowledge the uncertainty involved in the extrapolation of laboratory-derived anammox activities to the natural environment, the rates of net ammonium and nitrate consumption calculated as a function of depth indicate that nitrate reduction by denitrifiers coupled to anammox accounts for a substantial loss of

fixed inorganic nitrogen. In fact, the downward flux of nitrate ($\sim 7 \mu\text{mol m}^{-2} \text{ h}^{-1}$) is sufficient to oxidize all the ammonium ($\sim 5 \mu\text{mol m}^{-2} \text{ h}^{-1}$) diffusing up into the suboxic zone. If we assume that the area ($3 \times 10^5 \text{ km}^2$) below the shelf break ($< 200 \text{ m}$)⁸ represents the total surface area of the suboxic zone, 0.3 Tg of fixed inorganic nitrogen per year might be lost through nitrate reduction coupled to anammox. For comparison, the annual primary production of phytoplankton in the whole basin is $\sim 80 \text{ Tg}$ carbon (ref. 8), which is equivalent to 14 Tg of fixed organic N if we assume an atomic C/N ratio of 6.6 for phytoplankton¹⁷. Because more than 95% of this phytoplanktonic organic nitrogen is recycled in the upper 80 m (ref. 18), anammox might consume more than 40% of the fixed nitrogen that sinks down into the anoxic Black Sea water.

Moreover, these results demonstrate that anammox bacteria are abundant and are important in the nitrogen cycle of the Black Sea. In fact, the widespread occurrence of ammonium consumption in suboxic marine waters as well as in sediments⁷ suggests that anammox bacteria could have an important but as yet neglected role in the oceanic loss of fixed nitrogen. □

Methods

Nutrient analyses

Water samples for nutrient analyses were obtained by a pumpcast conductivity–temperature–depth (CTD) system equipped with an oxygen sensor. Before analyses, ZnCl_2 was added to the samples from the anoxic part of the water column to precipitate sulphide. Nitrate, nitrite and ammonium concentrations (detection limits 0.1, 0.01 and $0.5 \mu\text{M}$, respectively) were determined on board with an autoanalyser, immediately after sampling.

^{15}N incubations and analysis

Black Sea water collected from specific water depths was flushed for 1 h with argon and, after the addition of $500 \mu\text{M } ^{15}\text{NH}_4\text{Cl}$ and $100 \mu\text{M Na } ^{14}\text{NO}_2^-$, incubated for 4 days at *in situ* temperatures ($\sim 8^\circ\text{C}$). Subsequently, the samples were stored at 4°C until analysis. $^{14}\text{N}^{15}\text{N}$: $^{14}\text{N}^{14}\text{N}$ ratios were determined by gas chromatography–isotope ratio mass spectrometry and expressed as $\delta^{14}\text{N}^{15}\text{N}$ values ($\delta^{14}\text{N}^{15}\text{N} = [({}^{14}\text{N}^{15}\text{N}:{}^{14}\text{N}^{14}\text{N})_{\text{sample}} / ({}^{14}\text{N}^{15}\text{N}:{}^{14}\text{N}^{14}\text{N})_{\text{standard}}] - 1$; air was used as the standard).

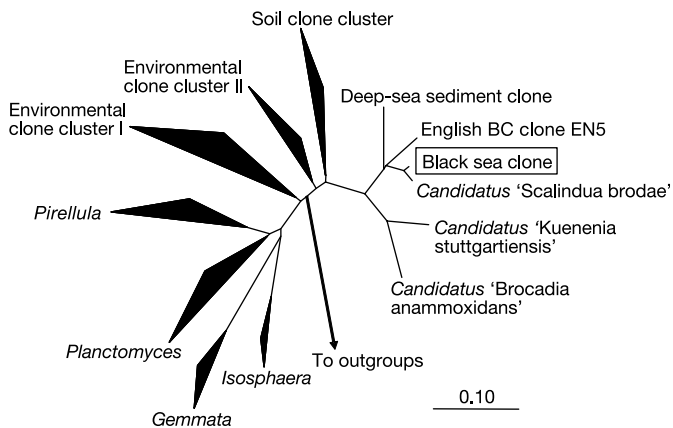


Figure 3 Phylogenetic tree of 16S rRNA gene sequences showing the order Planctomycetales and the position of the anammox-affiliated organisms from the Black Sea (indicated by a rectangle). The black triangles indicate phylogenetic groups. The bar represents 10% estimated sequence divergence. Deep-sea sediment clone is from ref. 26; English BC clone EN5 and *Candidatus 'Scalindua brodae'* are from M. Schmid, K. Walsh, R. Webb, W. I. Rijpstra, K. T. van de Pas Schoonen, T. C. J. Hill, B. F. Moffett, J. A. Fuerst, J.S.S.D., J. A. Harris, P. J. Shaw, M.S.M.J. and M. Strauss (unpublished observations).

Lipid analysis

Particulate organic matter for lipid analyses was collected from specific water depths by *in situ* filtration of large volumes (~1,000 l) of water through 292-mm diameter precombusted (at 450 °C) glass fibre filters (GFF; nominal pore size 0.7 µm) with *in situ* pumps. Because filtration through 0.7-µm pore-size filters could lead to an undersampling of anammox cells, the calculated ladderane lipid concentrations are minimum values. The GFF were extracted for 24 h in a Soxhlet apparatus to obtain the total lipid extracts. Aliquots of the total extracts were saponified after addition of an internal standard and separated into fatty acid and neutral lipid fractions. The fatty acid fractions were methylated and the neutral fractions were silylated and analysed by gas chromatography–mass spectrometry for the identification and quantification of ladderane lipids. Repeated concentration measurements were within ±10%.

Molecular cloning and phylogeny

DNA extraction, isolation and cloning were performed as described previously¹⁹. Phylogenetic analysis was performed with the ARB software package¹⁵. The phylogenetic tree is based on a maximum-likelihood analysis of different data sets.

FISH and microscopy

Filter material was stained with an oligonucleotide probe specific for Planctomycetes (Pla46, S-P-Planc-0046-a-A-18)²⁰, a newly designed Anammox probe (AmxBS820, S-⁺-BS-820-a-A-22 (5'-TAATTCCTCTACTTAGTGCCC-3')), a eubacterial probe (EUB338, S-D-Bact-0338-a-A-18)²¹ and DAPI to determine the abundance of anammox and total bacteria. FISH and DAPI staining were performed as described²² and the average number of anammox bacteria was determined by analysing 20 different slides.

Flux calculations

Nitrate and ammonium fluxes and ammonium oxidation rates were calculated from the concentration profiles and a vertical diffusion coefficient (K_z) with the program Profile²³. Published estimates of the vertical diffusion coefficient for the suboxic zone vary over an order of magnitude (0.02–0.7 cm² s⁻¹)^{8,24,25}. However, most calculations of chemical fluxes^{24,25} have used a K_z value close to the lower end of the range. Accordingly, a K_z of 0.04 cm² s⁻¹ was used here. The model predicted zones of net ammonium and nitrate consumption at 106–93 and 88–94 m, respectively.

Received 8 November 2002; accepted 3 February 2003; doi:10.1038/nature01472.

- Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO₂ in the ocean. *Nature* **387**, 272–275 (1997).
- Codispoti, L. A. & Christensen, J. P. Nitrification, denitrification and nitrous oxide cycling in the eastern tropical South Pacific Ocean. *Mar. Chem.* **16**, 277–300 (1985).
- Strous, M. *et al.* Missing lithotroph identified as new planctomycete. *Nature* **400**, 446–449 (1999).
- Sinninghe Damsté, J. S. *et al.* Linearly concatenated cyclobutane (ladderane) lipids from a dense bacterial membrane. *Nature* **419**, 708–712 (2002).
- Bender, M. *et al.* Organic carbon oxidation and benthic nitrogen and silica dynamics in San Clemente Basin, a continental borderland site. *Geochim. Cosmochim. Acta* **53**, 685–697 (1989).
- Codispoti, L. A., Friederich, G. E., Murray, J. W. & Sakamoto, C. M. Chemical variability in the Black Sea: implications of continuous vertical profiles that penetrated the oxic/anoxic interface. *Deep-Sea Res.* **38**, S691–S710 (1991).
- Thamdrup, B. & Dalsgaard, T. Production of N₂ through anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments. *Appl. Environ. Microbiol.* **68**, 1312–1318 (2002).
- Sorokin, Y. I. *The Black Sea—Ecology and Oceanography* (ed. Martens, K.) 1–875 (Backhuys, Leiden, 2002).
- Froelich, P. N. *et al.* Early oxidation of organic matter in pelagic sediments of the eastern equatorial Atlantic: Suboxic diagenesis. *Geochim. Cosmochim. Acta* **43**, 1075–1090 (1979).
- Richards, F. A. in *Chemical Oceanography* (eds Ripley, J. P. & Skirrow, G.) 611–645 (Academic, London, 1965).
- Murray, J. W., Codispoti, L. A. & Friederich, G. E. in *Aquatic Chemistry* (eds Huang, C. P., O'Melia, C. R. & Morgan, J. J.) 157–176 (American Chemical Society, Washington DC, 1995).
- Kuenen, J. G. & Jetten, M. S. M. Extraordinary anaerobic ammonium-oxidizing bacteria. *ASM News* **67**, 456–463 (2001).
- Dalsgaard, T. & Thamdrup, B. Factors controlling anaerobic ammonium oxidation with nitrite in marine sediments. *Appl. Environ. Microbiol.* **68**, 3802–3808 (2002).
- Mulder, A., van de Graaf, A. A., Robertson, L. A. & Kuenen, J. G. Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. *FEMS Microbiol. Ecol.* **16**, 177–184 (1995).
- Schmid, M. *et al.* Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. *Syst. Appl. Microbiol.* **23**, 93–106 (2000).
- Ward, B. B. & Kilpatrick, K. A. in *Black Sea Oceanography* (eds Izard, E. & Murray, J. W.) 111–124 (Kluwer Academic, Dordrecht, 1991).
- Redfield, A. C., Ketchum, B. H. & Richards, F. A. in *The Sea* (ed. Hill, M. N.) 26–77 (Interscience, New York, 1963).
- Karl, D. M. & Knauer, G. A. Microbial production and particle flux in the upper 350 m of the Black Sea. *Deep-Sea Res.* **38**, S921–S942 (1991).
- Schmid, M., Schmitz-Esser, S., Jetten, M. & Wagner, M. 16S–23S rDNA intergenic spacer and 23S rDNA of anaerobic ammonium-oxidizing bacteria: implications for phylogeny and *in situ* detection. *Environ. Microbiol.* **3**, 450–459 (2001).
- Neef, A., Amann, R., Schlesner, H. & Schleifer, K. H. Monitoring a widespread bacterial group: *in situ* detection of planctomycetes with 16S rRNA-targeted probes. *Microbiology* **144**, 3257–3266 (1998).
- Amann, R. I. *et al.* Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **6**, 1919–1925 (1990).
- Pernthaler, J., Glöckner, F. O., Schönhuber, W. & Amann, R. in *Methods in Microbiology* (ed. Paul, J. H.) 207–226 (Academic, London, 2001).
- Berg, P., Risgaard-Petersen, N. & Rysgaard, S. Interpretation of measured concentration profiles in

sediment pore water. *Limnol. Oceanogr.* **43**, 1500–1510 (1998).

- Oguz, T., Murray, J. W. & Callahan, A. E. Modeling redox cycling across the suboxic-anoxic interface zone in the Black Sea. *Deep-Sea Res.* **1** **48**, 761–787 (2001).
- Lewis, B. & Landing, W. M. The biogeochemistry of manganese and iron in the Black Sea. *Deep-Sea Res.* **38** (suppl. 2), S773–S803 (1991).
- Li, L., Kato, C. & Horikoshi, K. Bacterial diversity in deep-sea sediments from different depths. *Biodivers. Conserv.* **8**, 659–677 (1999).

Acknowledgements We thank L. Neretin for discussions; the Romanian and Turkish authorities for access to their national waters; the crew of the R/V Meteor for collaboration; S. Krüger, F. Pollehne and T. Leipe (IOW, Warnemünde) for operating the pumpcast and providing the CTD data; J. Eygensteyn, A. Pol, H. op den Camp, K. van de Pas Schoonen and G. Klockgether for analytical assistance. C. Hanfland and the AWI (Bremerhaven) provided the *in situ* pumps. The investigations were supported by the MPG, the University of Nijmegen, the TU Delft and the DFG. M.M.M.K. was financially supported by the EC Human Potential Programme Research Training Networks Activity (CT-net); A.O.S. was supported by a grant of the ALW; M. Schmid was supported by an EU grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.M.M.K. (e-mail: mkuypers@mpi-bremen.de).

Catastrophic ape decline in western equatorial Africa

Peter D. Walsh*, Kate A. Abernethy†‡, Magdalena Bermejo§, Rene Beyers||, Pauwel De Wachter¶, Marc Ella Akou¶, Bas Huijbregts¶, Daniel Idiata Mambounga#, Andre Kamdem Toham¶, Annelisa M. Kilbourn||, Sally A. Lahm☆, Stefanie Latour||, Fiona Maisels||**, Christian Mbina||, Yves Mihindou||, Sosthène Ndong Obiang#, Ernestine Ntsame Effa#, Malcolm P. Starkey||††, Paul Telfer†‡‡, Marc Thibault¶, Caroline E. G. Tutin†‡, Lee J. T. White|| & David S. Wilkie||

* Department of Ecology and Evolutionary Biology, Guyot Hall, Princeton, New Jersey 08540, USA

† Centre International de Recherches Médicales, BP 769, Franceville, Gabon

‡ Department of Biological and Molecular Sciences, University of Stirling, Stirling FK9 4LA, UK

§ Departamento Biología Animal (Vertebrados), Facultad de Biología, Universidad de Barcelona, Avda. Diagonal 645, 08028 Barcelona, Spain

|| Wildlife Conservation Society, Bronx, New York, New York 10460-1099, USA

¶ WWF Central Africa Regional Program Office, BP 9144, Libreville, Gabon

Ministère de l'Economie Forestière, des Eaux, de la Pêche chargé de l'Environnement et de la Protection de la Nature, Direction de la Faune et de la Chasse, BP 1128, Libreville, Gabon

☆ Institut de Recherche en Ecologie Tropicale, BP 13354, Libreville, Gabon

** Institute of Cell, Animal and Population Biology, Edinburgh University, Edinburgh EH9 3JT, UK

†† Department of Geography, University of Cambridge, Downing Place, Cambridge CB2 3EN, UK

‡‡ New York University, Department of Anthropology, 25 Waverly Place, New York, New York 10003, USA

Because rapidly expanding human populations have devastated gorilla (*Gorilla gorilla*) and common chimpanzee (*Pan troglodytes*) habitats in East and West Africa, the relatively intact forests of western equatorial Africa have been viewed as the last stronghold of African apes¹. Gabon and the Republic of Congo alone are thought to hold roughly 80% of the world's gorillas² and most of the common chimpanzees¹. Here we present survey results conservatively indicating that ape populations in Gabon declined by more than half between 1983 and 2000. The primary cause of the decline in ape numbers during this period was commercial hunting, facilitated by the rapid expansion of

mechanized logging. Furthermore, Ebola haemorrhagic fever is currently spreading through ape populations in Gabon and Congo and now rivals hunting as a threat to apes. Gorillas and common chimpanzees should be elevated immediately to 'critically endangered' status. Without aggressive investments in law enforcement, protected area management and Ebola prevention, the next decade will see our closest relatives pushed to the brink of extinction.

The plight of apes in western equatorial Africa (WEA; Fig. 1) has not been appreciated widely because previous assessments have assumed that the amount of intact forest cover is the primary determinant of ape abundance (Congo and Gabon retain about 60% and 80% of original forest cover, respectively^{3,4}). A widely cited 1995 estimate assuming moderate ape densities in all forested habitat put gorilla abundance in WEA at 110,000 (ref. 2). What this estimate did not take into account was the impact of threats other than deforestation. Rather than assuming that all forest held apes at some 'typical' density, from 1998 to 2002 we conducted surveys of ape sleeping nests at sites spread widely across Gabon. We compared our results with data from a national survey conducted in 1981–1983 (ref. 5), when high-density ape populations spanned Gabon (Fig. 2a). Our recent surveys suggest that high-density ape populations are now restricted to the southwest and northeast of the country (Fig. 2b). By far the best predictor of ape distribution in both time periods was distance from the nearest of Gabon's four major urban centres, which explained 61% of the variance in nest group encounter rate (Fig. 3). The effect of cities was already evident in 1983, but ape distribution has since contracted markedly, with healthy ape populations found only in very remote areas. The other good predictor of encounter rate of nest groups was distance from the nearest documented human Ebola outbreak site⁶ ($R^2 = 0.63$ for a model including both variables, $P = 0.02$, $n = 89$). When data from the two time periods are compared, they indicate an ape decline of 56% (95% confidence interval = 35–70%). This figure may underestimate substantially the true magnitude of ape decline (see Methods).

Sharp declines in ape abundance are almost certainly not restricted to Gabon. Neighbouring Congo has a higher rural

human density (<http://apps.fao.org>) and deforestation rate⁴ than Gabon, and apes in northwest Congo are in the midst of a devastating Ebola epidemic (<http://www.who.int>). The situations in other WEA countries that still hold appreciable ape populations (that is, Cameroon, Central African Republic and Equatorial Guinea¹) are probably worse as a consequence of their higher human densities (<http://apps.fao.org>) and deforestation rates⁴. Also of great concern is the suspected but largely undocumented crash of ape populations in eastern Democratic Republic of Congo, the consequence of hunting spurred by recent civil unrest.

The World Conservation Union (IUCN) red list of threatened species now classes both *G. gorilla* and *P. troglodytes* as 'endangered' (<http://www.redlist.org>). Our results suggest that there should be an immediate reclassification of both species to 'critically endangered' status. Under criterion CR-A2, a species should be considered critically endangered if it is expected to suffer "A reduction of at least 80% ... within the next 10 years or three generations." Assuming a constant decline rate between the end of the early surveys (1983) and the midpoint of recent surveys (2000), the annual decline rate is 4.7%. At this rate, ape populations would decline by an additional 80% within 33 years: a generation and one half for chimpanzees and perhaps two generations for gorillas (Fig. 4a). Given that our decline estimates are conservative and that decline rates have probably been accelerating during recent years, the 80% threshold will probably be reached much sooner. Even if decline can be halted, ape demographic rates are so low that population recovery will be very slow (Fig. 4b).

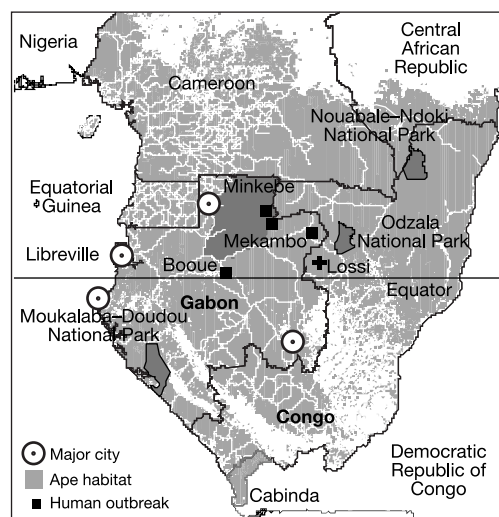


Figure 1 Western equatorial Africa. Commercial hunting is concentrated near transport routes (major roads in white), whereas Ebola impact has been heaviest in remote areas. Ape density in the Minkébé forest block (shaded area in northern Gabon) has dropped by about 99% over the past decade. The Ebola epidemic is currently most severe between Lossi and Odzala National Park. Observations of unpredated ape carcasses span 600 km, from 40 km west of Booue to the Nouabale–Ndoki region of northern Congo.

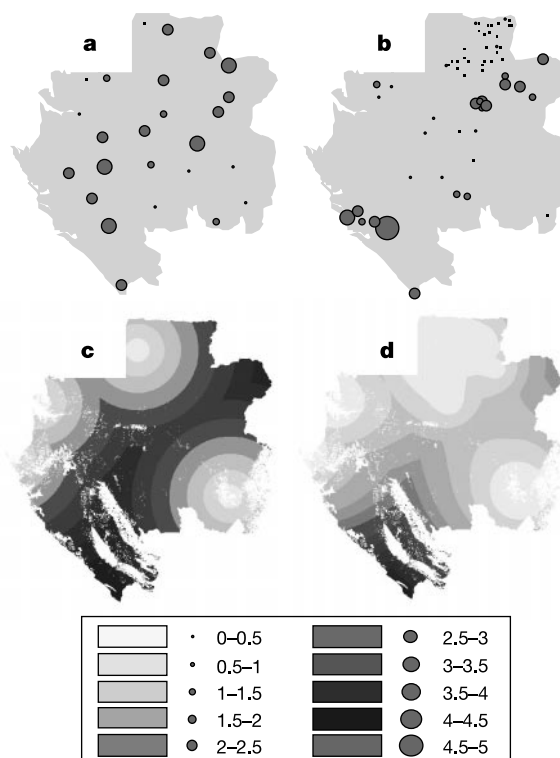


Figure 2 Ape nest encounter rates in Gabon. **a, b**, Encounter rates of ape nest groups (nest groups per km) from early 1980s (**a**) and recent surveys (**b**). The smallest circles include sampling sites with zero nest group encounters. **c, d**, Individual nest encounter rates (nests per km) predicted by distance from nearest major urban centre and distance from nearest human Ebola outbreak (see Methods): **c**, early 1980s; **d**, recent surveys. Predictions are corrected to reflect positive correlation between nest group size and distance from nearest human Ebola outbreak. Regression surfaces do not accurately represent smaller-scale variation in encounter rate.

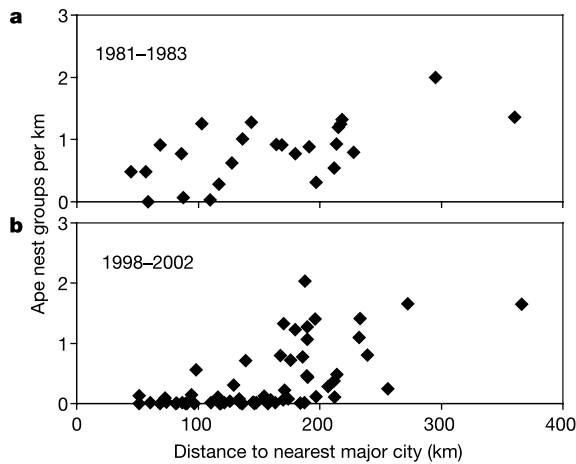


Figure 3 Encounter rate of ape nest groups plotted as a function of distance to the closest of Gabon's four major urban areas (anticlockwise from Libreville are Port Gentil, Franceville and Oyem).

Over the past decade, the intensity and spatial extent of commercial hunting have increased greatly—driven by the rapid expansion of mechanized logging and by economic migration and resettlement policies^{7,8}, which have concentrated salaried bushmeat customers in towns and cities (<http://apps.fao.org>)⁹. The transportation infrastructure provided by logging has transformed hunting from a primarily subsistence activity into a commercial enterprise. Villages near logging roads hunt much more intensively than villages that lack market access⁹. Furthermore, hunting is no longer confined to local villagers or to the immediate proximity of villages. Organized groups of hunters use logging roads and vehicles to penetrate deep into remote areas (including parks and wildlife reserves)¹⁰, then export bushmeat to nearby logging towns (where logging employees eat more bushmeat than local villagers⁹), regional transportation hubs, and even large cities hundreds of kilometres away¹¹. For the four market towns where we observed gorilla sales, time of travel from the Gabonese capital Libreville was an excellent predictor of bushmeat price per kg ($n = 31$ species, $F = 67.86$, $P < 0.001$). Bushmeat costs about one-third as much as alternative sources of protein (frozen chicken, beef and fish) in remote villages, but 50% more in Libreville ($R^2 = 0.91$, $P = 0.048$), where it is consumed on only about 2% of days. Most ape hunting in Gabon is probably not driven by a specific consumer demand for gorilla or chimpanzee meat. Forty-one (42%) out of 98 people that we interviewed in Libreville cited gorilla as a meat that they did not eat. Now that bushmeat has become too expensive for poor Librevillois, smaller regional cities and towns have probably become the principal centres of urban bushmeat consumption. But even in rural areas, ape meat does not make a substantive contribution to food security: maximum sustainable yield of chimpanzee meat is 10–50-fold lower than that of preferred market species (Supplementary Information).

Since 1994, Gabon has had four well-documented human Ebola epidemics, all in the northeast. Ape carcasses have been found near the sites of three of the human epidemics, with apes testing positive for Ebola in two cases⁶ (<http://www.who.int>). The impact of Ebola on ape populations has been most extensive in the Minkébé region, a lightly settled, roadless block of about 32,000 km² (Fig. 1)¹². Minkébé showed moderate⁵ to high¹³ encounter rates of ape nest groups in pre-Ebola surveys, including a 1991 study in which 75 ape nest groups were observed in only 20 km of line transects, a rate higher than all but one published report from WEA¹⁴. By 2000, nest group encounter rates on reconnaissance surveys (2,700 km) and line transects (17 km) yielded only 91 nest groups, a drop of about

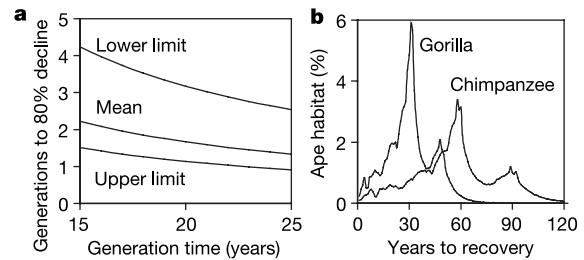


Figure 4 Decline and recovery projections. **a**, Generations until Gabonese apes decline by an additional 80%, assuming the mean decline rate observed between 1983–2000 (4.7%), and upper and lower 95% confidence limits for that decline rate. Chimpanzee generation time is about 22 years, whereas gorilla generation time is slightly shorter. **b**, Frequency distribution of time necessary for gorilla and chimpanzee populations, across Gabon to recover to 99% of 1982 levels. Projections are based on ideal conditions, including no hunting, no Ebola, no density-dependent population regulation, and highly optimistic birth rates (see Supplementary Information for details).

99%. Hunting was not a probable cause of this marked decline, but hunters described finding large numbers of dead apes both near human epidemic sites and in other sectors of Minkébé¹². Other remote areas of central and northeastern Gabon covered in our surveys also showed little sign of hunting pressure (for example, they had high encounter rates of elephant dung comparable to rates from the late 1980s¹⁵), but exhibited unexpectedly low ape densities.

The ape Ebola epidemic is gradually working its way south and west. Over the past 18 months, wildlife conservation workers and an international research team have found many ape carcasses in Congo, across the border from the 2001 human epidemic site at Mekambo (<http://www.who.int>). Since December, the carcasses of seven gorillas from a study population at Lossi have been found, along with five chimpanzee carcasses. The study population originally included 143 individually identified gorillas, but after exhaustive searches only seven have been found alive. The epidemic is now approaching Odzala National Park, which boasts the world's highest recorded gorilla and chimpanzee densities¹⁶.

The precipitous decline of apes in WEA is a major conservation crisis. Responding effectively to the Ebola epidemic will require intensified research on reservoir and host dynamics, with a more quantitative, ecological emphasis. Research on vaccines and modes of deployment is also a high priority. Other interventions in the Ebola transmission chain should be investigated actively. The solution to the hunting crisis will not result from investment in poverty alleviation or sustainable development. Rural poverty is too intractable to be solved before hunting further demolishes ape populations, and the social and economic contexts in WEA are such that alternative use strategies such as gorilla tourism cannot provide national governments with enough income to offset the large opportunity costs of forgoing extractive use of resources^{17,18}. The bulk of conservation investment should, instead, be focused on formally protected areas that still have enough apes to be viable in the long run. The immediate priority is a massive investment in law enforcement, which has long been underfunded in WEA. The long-term priority should be building national capacity to conduct all aspects of protected area management. In addition, logging and oil companies should be compelled to control illegal bushmeat hunting, transport and consumption on their concessions. Funds to pay for hunting and Ebola control efforts must be generated in developed countries where apes are cherished as a vital element of our natural heritage. We must also provide national governments with economic incentives, linking aid and debt relief to verifiable measures of conservation performance. The stark truth is that if we do not act decisively our children may live in a world without wild apes. □

Methods

Surveys

Our data for the recent survey period come from 261 km of formal line transects¹⁹ and 4,793 km of reconnaissance surveys ('recces'²⁰) performed while evaluating existing and potential protected areas in Gabon. The sampling sites were chosen because they were thought to contain uncharacteristically high large-mammal densities, whereas sampling locations from the national survey in 1981–1983 were spread more uniformly across Gabon. The survey data were also collected before the ongoing Ebola epidemic in northeast Gabon and neighbouring Congo (<http://www.who.int>). Therefore, the figures for decline that we report below are likely to underestimate substantially the true magnitude of ape decline.

Surveys in both sampling time periods were based on counts of sleeping nests rather than live gorillas or chimpanzees, which are secretive and difficult to observe during surveys. Because discriminating between gorilla and chimpanzee nests can be difficult, we pooled gorilla and chimpanzee data and analysed the rate at which ape nest groups (not individual nests) were encountered along line transects and recces. Raw encounter rate data are shown in Fig. 2. Evidence that encounter rates of ape nest groups should be directly proportional to ape density is provided in Supplementary Information, along with a detailed discussion of analytical methods.

Measurement of ape decline

To estimate the magnitude of ape decline, we took a 'spatial modelling' approach²¹. For each sampling period, we estimated the functional relationship between nest group encounter rate and important predictor variables, cut Gabon into a 30 arcsec grid (approximately 1 km² cells), and interpolated encounter rate in each grid cell given the predictor variable values for that cell. We confined ourselves to analysing the effects of two predictor variables: distance from the closest of Gabon's major urban centres and distance from documented human Ebola outbreaks. We do not include analyses with predictor variables that varied on a smaller scale because our sampling sites tended not to include the full range of predictor values on smaller scales. For example, few sampling sites were near towns or heavily deforested areas, so that analyses of the effects of these factors on encounter rates of nest groups tended to produce obviously spurious results. Our analyses indicated a strong increase in nest group size with increasing distance from human Ebola outbreak sites ($R^2 = 0.012$, $P = 0.00002$, $n = 1,434$). Therefore, when interpolating the value for each cell in the Gabon grid, we multiplied the predicted nest group encounter rate by the predicted nest group size. We then estimated the percentage ape decline by averaging nest encounter rates across grid cells, dividing the average for the recent surveys by the 1981–1983 average, subtracting the quotient from one, and multiplying by 100. Without including the effect of declining group size, the analysis suggested a nest group encounter rate decline of 46% (95% confidence interval = 23–63%).

Analysis of bushmeat sale and consumption

Sales of bushmeat were recorded at 11 markets in Gabon between January 2000 and July 2002. Analysed surveys included between 21 and 426 days per market and 80,528 bushmeat sales from 42 species, of which 7,686 (9.5%) were of whole animals. Prices per unit mass for whole animal sales were calculated using published average adult body weights²². Prices of alternative protein sources were recorded monthly in a sample of shops in each town. In Libreville, we visited 518 randomly chosen households once each between January and April 2001 for a total of 1,793 consumption days. We asked about: (1) consumption of meat over the previous three days; (2) household income in the previous month; and (3) household assets.

Received 18 December 2002; accepted 17 March 2003; doi:10.1038/nature01566.

Published online 6 April 2003.

- Oates, J. F. *African Primates: Status Survey & Action Plan* (IUCN, Cambridge, 1996).
- Harcourt, A. H. Is the gorilla a threatened species? How should we judge? *Biol. Conserv.* **75**, 165–176 (1996).
- Minnemeyer, S., Walker, T., Collomb, J. G., Cotton, L. & Bryant, D. *An Analysis of Access to Central Africa's Rainforests* (World Resources Institute, Washington, 2002).
- Roberts, L. (ed.) *World Resources 292* (Oxford Univ. Press, New York, 1998).
- Tutin, C. E. G. & Fernandez, M. Nationwide census of gorilla (*Gorilla g. gorilla*) and chimpanzee (*Pan t. troglodytes*) populations in Gabon. *Am. J. Primatol.* **6**, 313–336 (1984).
- Georges, A. J. et al. Ebola hemorrhagic fever outbreaks in Gabon, 1994–1997: Epidemiologic and health control issues. *J. Infect. Dis.* **179**, S65–S75 (1999).
- Chamberlin, C. Migration of Fang into central Gabon during the 19th century—a new interpretation. *Int. J. Afr. Hist. Stud.* **11**, 429–456 (1978).
- Pourtier, R. *Le Gabon* Vol. 1 (Harmattan, Paris, 1989).
- Wilkie, D. S., Shaw, E., Rotberg, F., Morelli, G. & Auzel, P. Roads, development, and conservation in the Congo Basin. *Conserv. Biol.* **14**, 1614–1622 (2000).
- Rapport des activités de la Lutte Antibrucellose, Brigade de Faune de la Lope (Direction de la Faune et de la Chasse, Libreville, Gabon, 1999).
- Angoué, C. et al. in *L'avenir des peuples des forêts tropicales—Rapport final* Vol. III (eds Bahuchet, S. & de Maret, P.) (Université Libre de Bruxelles, Brussels, 2000).
- Huysbregts, B., de Wachter, P. & Ndong Obiang, L. S. Ebola and the decline of gorilla (*Gorilla gorilla*) and chimpanzee (*Pan troglodytes*) populations in the Minkebe Forest, north-eastern Gabon. *Oryx* (in the press).
- Lahm, S. A. *Ecology and Economics of Human/Wildlife Interaction in Northeastern Gabon*. Dissertation, New York Univ. (1993).
- Brugière, D. & Sakom, D. Population density and nesting behaviour of lowland gorillas (*Gorilla gorilla gorilla*) in the Ngotto forest, Central African Republic. *J. Zool.* **255**, 251–259 (2001).
- Barnes, R. F. W., Barnes, K. L., Alers, M. P. T. & Blom, A. Man determines the distribution of elephants in the rain forests of northeastern Gabon. *Afr. J. Ecol.* **29**, 54–63 (1991).

- Bermejo, M. Status and conservation of primates in Odzala National Park, Republic of the Congo. *Oryx* **33**, 323–331 (1999).
- Blom, A. The monetary impact of tourism on protected area management and the local economy in Dzanga-Sangha (Central African Republic). *J. Sust. Tourism* **8**, 175–189 (2001).
- Wilkie, D. S. & Carpenter, J. Can nature tourism help finance protected areas in the Congo Basin? *Oryx* **33**, 333–339 (1999).
- Buckland, S. T., Anderson, D. R., Burnham, K. P. & Laake, J. L. *Distance Sampling: Estimating Abundance of Biological Populations* (Chapman & Hall, London, 1993).
- Walsh, P. D. & White, L. J. T. What it will take to monitor forest elephant populations. *Conserv. Biol.* **13**, 1194–1202 (1999).
- Hedley, S. L., Buckland, S. T. & Borchers, D. L. Spatial modelling from line transect data. *J. Cetacean Res. Manage.* **1**, 255–264 (1999).
- Kingdon, J. *The Kingdon Field Guide to African Mammals* (Academic, San Diego, 1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements This paper is dedicated to Annelisa Kilbourn who died on 2 November 2002 in a plane crash en route to her Ebola research site. For support of ape survey work we thank USAID-CARPE, the DGIS, the MacArthur Foundation, WCS, WWF and the République Gabonaise. Funding for the socioeconomic surveys was provided by NSF D. Purves, S. Sandin and J. Regetz provided helpful comments on statistical analyses.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.D.W. (e-mail: pwalsch@princeton.edu).

Subsecond dopamine release promotes cocaine seeking

Paul E. M. Phillips^{*†‡}, Garret D. Stuber^{‡§}, Michael L. A. V. Heien[†], R. Mark Wightman^{†‡§} & Regina M. Carelli^{*§}

^{*} Department of Psychology, [†] Department of Chemistry, [‡] Neuroscience Center, and [§] Curriculum in Neurobiology, University of North Carolina, Chapel Hill, North Carolina 27599, USA

The dopamine-containing projection from the ventral tegmental area of the midbrain to the nucleus accumbens is critically involved in mediating the reinforcing properties of cocaine^{1,2}. Although neurons in this area respond to rewards on a subsecond timescale^{3,4}, neurochemical studies have only addressed the role of dopamine in drug addiction by examining changes in the tonic (minute-to-minute) levels of extracellular dopamine^{5–9}. To investigate the role of phasic (subsecond) dopamine signalling¹⁰, we measured dopamine every 100 ms in the nucleus accumbens using electrochemical technology¹¹. Rapid changes in extracellular dopamine concentration were observed at key aspects of drug-taking behaviour in rats. Before lever presses for cocaine, there was an increase in dopamine that coincided with the initiation of drug-seeking behaviours. Notably, these behaviours could be reproduced by electrically evoking dopamine release on this timescale. After lever presses, there were further increases in dopamine concentration at the concurrent presentation of cocaine-related cues. These cues alone also elicited similar, rapid dopamine signalling, but only in animals where they had previously been paired to cocaine delivery. These findings reveal an unprecedented role for dopamine in the regulation of drug taking in real time.

Dopamine-containing neurons in the ventral tegmental area are synchronously activated (55–80% of neurons burst for up to 200 ms) on presentation of natural reinforcers or associated cues³. It has been proposed that this burst of activity causes transient changes in forebrain extracellular dopamine that provide a real-time learning signal for reward¹². Such a dopamine surge may

modulate specific aspects of goal-directed behaviours for natural rewards and perhaps drug rewards, and may have evolved through the formation of learned associations between these reinforcers and environmental cues. To investigate these possibilities with respect to drug abuse we used fast-scan cyclic voltammetry¹¹ to measure subsecond changes in dopamine in the nucleus accumbens of rats performing cocaine self-administration.

A carbon-fibre microelectrode was positioned into the nucleus accumbens of rats that were trained to lever press for intravenous cocaine, and extracellular dopamine was monitored every 100 ms. Release of dopamine was evoked before the session by electrical stimulation (24 pulses, 60 Hz, 120 μ A) of the ventral tegmental area (Fig. 1) to confirm that the microelectrode was in a dopamine-rich area. Electrical stimulations were not applied again until immediately after the session, when they were used to verify that evoked dopamine release was still detectable. During experimental sessions, rats ($n = 6$) lever pressed for cocaine (0.33 mg) 3.4 ± 0.9 times in quick succession within the first few minutes (load-up), followed by stable rates of lever pressing (every 364 ± 20 s) for the remainder of the session (19.7 ± 1.5 total responses over 2 h; see Fig. 2a). Time locked to every operant response for cocaine (116 responses), there were detectable transient increases in extracellular dopamine that occurred both in the seconds before (pre-response) and after (post-response) the lever press (Fig. 2b, c).

The post-response component was tightly time locked to the lever press across trials and animals. There was a sharp inflection in extracellular dopamine, which peaked (64.9 ± 16.1 nM) at 1.8 ± 0.4 s and returned to baseline 4.7 ± 0.5 s after the response. This inflection occurred at the lever press and the concurrent onset of cocaine delivery paired to an audiovisual cue. In humans, cocaine

paraphernalia and other drug-related cues can elicit intense craving for cocaine and thereby influence drug taking^{13,14}. Furthermore, stimuli that had been paired with cocaine cause tonic increases in dopamine in the core (but not shell) of the nucleus accumbens in rats⁹. To test whether these types of cues might evoke phasic dopamine signalling, 'probe' trials ($n = 6$ rats) were completed where the audiovisual stimulus was randomly presented by the experimenter in the absence of a lever press or cocaine delivery. These caused rapid increases (93.9 ± 12.2 nM) in the extracellular dopamine concentration (Fig. 3), beginning at 0.1 ± 0.0 s, peaking at 2.2 ± 0.5 s and returning to baseline 5.7 ± 0.8 s from the probe onset. The concentration, onset latency, rise time and duration of these signals were not significantly different when the probes were presented before (4 trials) or during (12 trials) the self-administration session ($P > 0.05$), indicating that the presence of cocaine was not a requirement for the neurochemical change. The amplitude, time of peak and return to baseline after the cue onset were not significantly different than the post-response component for lever pressing at the same recording site ($P > 0.05$, paired t -test). Conversely, the audiovisual stimulus did not evoke dopamine release in rats where this stimulus had never been paired to cocaine delivery ($n = 4$), not even in the presence of cocaine ($n = 3$; Fig. 3b). Hence, this signalling requires learned associations between cocaine and stimuli that were previously paired with its infusion.

The pre-response component of dopamine signalling was variable in its synchronization to the lever press (and was therefore diminished in the signal-averaged trace). However, in the 4.1 ± 0.6 s before the lever press, there was consistently a rise in extracellular dopamine (Fig. 2c, second arrow) that was typically preceded with a more transient increase (7.7 ± 0.6 s before lever

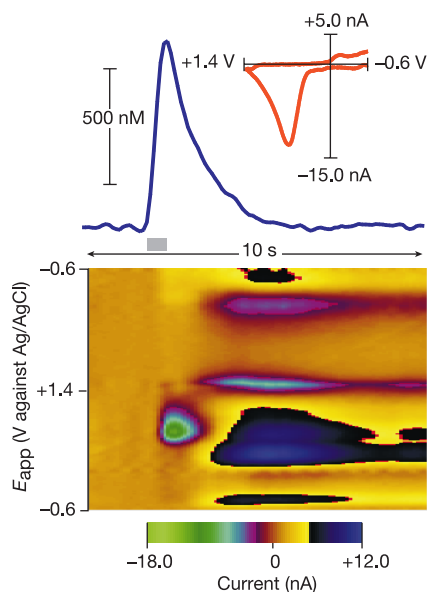


Figure 1 Dopamine release in the nucleus accumbens evoked by a stimulus train (24 pulses, 60 Hz, 120 μ A; represented by the grey bar). The top part of the figure shows the time course of dopamine concentration change. Inset, cyclic voltammogram obtained at the end of the stimulus that is identical to that for dopamine measured after the *in vivo* experiment. The bottom part of the figure shows a two-dimensional representation of the voltammetric data from a single animal. Consecutive cyclic voltammograms are plotted along the x-axis (time) with the applied potential (E_{app}) on the y-axis and current changes encoded in colour. Dopamine can be seen around the stimulation by the green peak (oxidation) in the lower part of the trace that is accompanied by a smaller yellow peak (reduction) in the upper part. The events that take place later in the trace (multiple peaks) are not due to changes in dopamine concentration, but a basic pH change in the extracellular space. This does not interfere with the dopamine signal because of the use of the differential measurement.

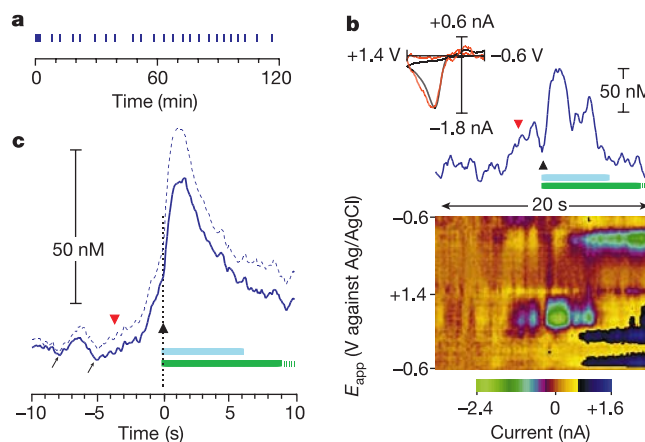


Figure 2 Rapid increase in extracellular dopamine in the nucleus accumbens relative to the lever-press response for cocaine. **a**, Individual lever-press responses (vertical lines) for a representative rat are shown against time in the session. **b**, The red, inverted triangle denotes the final lever approach before a lever press, represented by the black triangle, during this session. The light-blue bar represents cocaine infusion (0.33 mg, 6 s); the green bar represents the audiovisual stimulus (20 s), which is truncated at 10 s. The top part shows the time course of dopamine concentration change. Inset, the cyclic voltammogram obtained at the maximal dopamine change (red line) is superimposed on a current-normalized one obtained during electrical stimulation (black line; $r^2 = 0.93$). The bottom part shows a two-dimensional representation of the voltammetric data with current changes encoded in colour. Dopamine changes and the subsequent basic pH change are revealed. The pH change was removed from the dopamine signal using a differential measurement, and would otherwise erroneously appear as a drop in dopamine below the baseline. **c**, The solid blue line is the mean dopamine change across all animals ($n = 6$) around the lever press, and the dashed blue line is the mean plus standard error. Increases in dopamine before the lever press are highlighted by the arrows. The red, inverted triangle shows the mean time of initiation of the final approach to the lever before pressing (3.7 ± 0.5 s before press).

press; Fig. 2c, first arrow). During this period of time rats moved around the chamber to varying degrees, and then approached the lever before pressing. The initiation of this final approach (3.7 ± 0.5 s before the lever press) was after the first dopamine transient and during the rise in the second transient (Fig. 2c), and so we hypothesized that it is triggered by the dopamine changes. To test whether transient changes in extracellular dopamine were sufficient to initiate cocaine seeking, electrical stimulation of the ventral tegmental area (24 pulses, 60 Hz, 120 μ A) was used to deliver controlled, phasic dopamine transients to the nucleus accumbens. Stimulations were repeated every 120 s throughout sessions of cocaine self administration ($n = 6$) that were otherwise identical to those in the first experiment. These elicited rapid changes in extracellular dopamine in the nucleus accumbens (rise time of about 0.6 s). When the behaviour in these animals was examined, it was apparent that the electrical stimulations had a profound effect (animal group by time interaction: $F_{30,310} = 3.06$, $P < 0.0001$, two-way analysis of variance (ANOVA); Fig. 4a). Responding behaviour was highly elevated at 120 and 240 s after the previous lever press, with an apparent trend at higher multiples of 120 s. The amplitude of these successive peaks decreased in an exponential manner, indicative of an event every 120 s that had a fixed probability of triggering lever pressing. Clearly the lever-pressing behaviour had synchronized to the stimulations, as this pattern was not apparent in control (non-stimulated) rats.

To examine this synchronization with the stimulations more closely, the distribution of lever pressing was replotted relative to stimulation delivery (Fig. 4b). Predictably, the time-matched controls that were not stimulated at time zero showed a random distribution of lever presses with respect to this time. However, in animals that were stimulated, the temporal pattern of responding

for cocaine was significantly skewed (animal group by time interaction: $F_{23,240} = 2.88$, $P < 0.0001$, two-way ANOVA) with a maximal probability of lever pressing 5–15 s after a stimulation. The videotaped behavioural records revealed that stimulation was immediately followed by the behavioural sequence that normally preceded a lever press, including disengagement from stereotypy, 'superstitious' movement around the chamber and finally the approach to the lever before pressing. Notably, the delay between stimulation and lever press is in the range that the pre-response dopamine component precedes the lever press in non-stimulated rats (Fig. 2c). On the occasions where stimulation did not elicit a lever press, behaviour typically ranged from an immediate pause in stereotypy to all of the behaviours up to and including the lever approach. Thus, stimulation was effective in initiating behaviours, but the sequence was not always completed to yield cocaine delivery. Similarly, non-stimulated rats often approach the lever between presses without completing the response.

Our findings reveal that rapid dopamine transmission occurs during key components of cocaine-seeking behaviour and during presentation of cocaine-associated stimuli. Rather than a pharmacological effect, dopamine increases in response to cues that have a learned association with cocaine. Although previously suspected¹⁵, the rapid nature of these signals has precluded them from being detected until now, as previous studies used lower temporal resolution techniques (sampling 200–6,000 times less frequently¹⁰) that measure tonic dopamine levels. Nonetheless, tonic levels are of utmost relevance to the current findings because they regulate phasic dopamine release. Cocaine inhibits the dopamine transporter, slowing the clearance of extracellular dopamine, and thus increases tonic levels⁷. When dopamine is tonically high in the nucleus accumbens, terminal autoreceptors are activated, resulting in inhibition of phasic dopamine release¹⁶. In addition, cocaine elevates dopamine in the ventral tegmental area and thereby activates somatodendritic autoreceptors that inhibit dopamine-mediated cell firing^{17,18}, presumably reducing the occurrence of phasic events. During the intermittent pattern of cocaine delivery

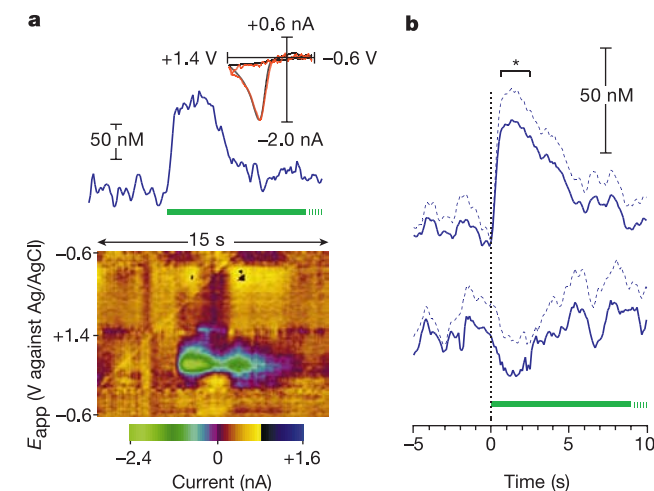


Figure 3 Rapid increase in extracellular dopamine after probe presentation of cocaine-associated stimuli. **a**, In rats trained to self administer cocaine paired with the audiovisual stimulus, the extracellular dopamine concentration transiently increases during presentation of this stimulus alone (green bar, truncated at 10 s). A representative time course of dopamine concentration change and its associated colour plot reveal that dopamine changes are apparent immediately after probe presentations. The cyclic voltammogram at the maximal dopamine change (red line) concurs to that during electrical stimulation (black line; $r^2 = 0.97$). **b**, The solid blue lines are the mean dopamine change around the 20-s stimulus presentation (green bar that starts at time 0, truncated at 10 s), and the dashed blue lines are the mean plus standard error. The upper trace is in animals that had been trained to lever press for cocaine paired to the stimulus ($n = 6$); the lower trace is in animals that had no experience of pairing between the stimulus and cocaine ($n = 3$). There was a significant interaction between animal group and time ($F_{150,1057} = 1.83$, $P < 0.0001$, two-way ANOVA). Asterisk, $P < 0.05$ compared with non-trained rats (Bonferroni post-tests).

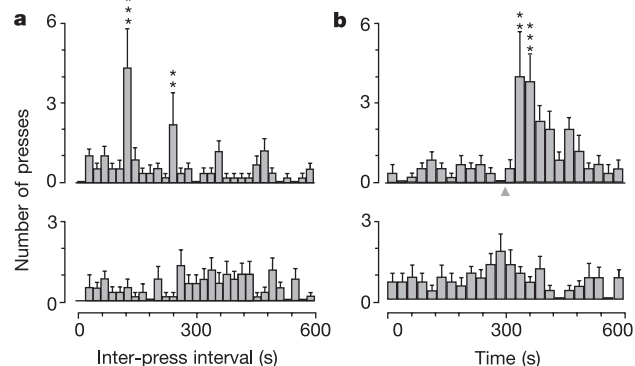


Figure 4 Effect of electrically evoked dopamine release on lever-press responding for cocaine. **a**, The histograms represent the distribution (20-s bins, truncated at 600 s) of intervals between lever presses in rats that received electrical stimulation (24 pulses, 60 Hz, 120 μ A every 120 s) to the ventral tegmental area (top, $n = 6$) and non-stimulated controls (bottom, $n = 6$). There was a significant interaction between animal group and time ($F_{30,310} = 3.06$, $P < 0.0001$, two-way ANOVA). Double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$ compared with non-stimulated controls (Bonferroni post-tests). **b**, The upper histogram represents the temporal distribution (5-s bins) of lever presses with respect to stimulation. The time of the stimulation was designated as zero (grey triangle). The lower histogram shows the distribution of lever presses over the same window for time-matched (non-stimulated) controls. There was a significant interaction between animal group and time ($F_{23,240} = 2.88$, $P < 0.0001$, two-way ANOVA). Double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$ compared with non-stimulated controls (Bonferroni post-tests).

that is inherent to self administration, tonic levels of dopamine vary. Extracellular dopamine accumulates during the load-up phase and then cycles, peaking a few minutes after each cocaine infusion and then falling. Once dopamine falls below a threshold concentration, rats lever press for further cocaine^{5,6}. It is at this time, when tonic regulation of phasic dopamine is minimal, that the dopamine transients accompanying cocaine seeking occur.

The finding that electrically evoked dopamine transients influence behaviour and can actually trigger lever-press responses for cocaine demonstrates that subsecond dopamine has a pivotal role in drug-seeking behaviours. It is important to note that the behavioural consequences of rapid dopamine transmission may be context specific. For instance, the increase in dopamine after each lever press (post-response) does not cause further cocaine seeking—the animal engages in stereotypy for about 6 min before pressing again. This is not altogether surprising as dopamine acts as a ‘neuromodulator’ of transmission in the nucleus accumbens, and so its effects are dependent on the activity of other afferents¹⁹ (which may alter with context²⁰). Specifically, dopamine-containing synapses are appositely located for modulation of information conveyed by descending glutamate-mediated pathways^{21,22}. This information relates to memory, drive and motivation and is integrated by the nucleus accumbens for the generation of goal-directed movement^{23,24}. Notably, the onset and duration of the rapid dopamine signals are similar to phasic changes in nucleus accumbens cell firing that occur under similar experimental conditions^{24–27}. We propose that by modulating neural activity in this structure, subsecond dopamine release has a key, real-time role in drug-seeking behaviour. □

Methods

Cocaine self administration

Male Sprague–Dawley rats were implanted with an indwelling jugular catheter and one week later were trained to self administer cocaine during 2-h daily sessions. The beginning of the session was signalled by extension of a lever into the chamber below an illuminated cue light. Each lever press resulted in cocaine delivery (0.33 mg, 6 s) and a 20-s audiovisual stimulus²⁵. This stimulus consisted of a change to general lighting of the chamber from the focal cue light and a continuous auditory tone. During this 20-s period lever pressing had no consequences. Once stable responding²⁵ had been achieved in three consecutive sessions, surgery for fast-scan cyclic voltammetry was performed²⁸. Approximately one week later, rats were allowed to self administer cocaine until stable behaviour was re-established (one to two sessions).

Fast-scan cyclic voltammetry

Dopamine was detected by oxidizing it with a carbon-fibre microelectrode using fast-scan cyclic voltammetry. The carbon-fibre microelectrode was held at -0.6 V against Ag/AgCl between scans and then periodically driven to $+1.4$ V and back in a triangular fashion at 400 V s⁻¹. We repeated scans every 100 ms. For analyte identification, current during a voltammetric scan was plotted against the applied potential to yield a cyclic voltammogram. Once dopamine had been voltammetrically verified, the current at its peak oxidation potential was plotted against time to reveal the temporal profile of dopamine concentration changes. To remove non-dopamine-mediated signals that occur at the potential of the dopamine oxidation peak (for example, from pH or movement artefacts), we used a differential measurement between the current at this potential (around $+0.70$ V against Ag/AgCl) and another one that included this interference (but not dopamine). This was converted to dopamine concentration by calibration of the electrode after *in vivo* use.

Non-trained rats

Rats ($n = 4$) with no experience of cocaine were surgically prepared for voltammetry. After recovery, they were habituated to the recording chamber and presented the audiovisual stimulus (5 times, about 5 min apart) on several daily sessions. On the experimental day, dopamine was monitored in the core of the nucleus accumbens and this model was repeated. Some rats ($n = 3$) were then administered cocaine (10 mg per kg intraperitoneally) and 10 min later the experiment was repeated once more.

Signal identification

Dopamine was identified with chemical, anatomical, physiological and pharmacological criteria²⁹. Anatomical evidence for dopamine was provided by post-mortem, histological verification. This confirmed that all the recording sites were in the core of the nucleus accumbens. By stimulating dopamine-containing cell bodies before and after behavioural sessions and detecting dopamine at the recording site (a well-characterized signal³⁰) we provided physiological evidence that we were recording from an area that could support rapid dopamine release. Chemical evidence for dopamine was provided by the cyclic

voltammogram, which offers information specific to the analyte. The cyclic voltammograms of signals after lever press responses or audiovisual stimuli were compared to those from electrical stimulations at the same recording site and those from *in vitro* calibration of the electrode. It is important to note that in addition to electro-oxidation, both movement artefacts and ionic changes in the extracellular space (especially pH; see Fig. 1) produce current at the electrode. These can be identified readily with fast-scan cyclic voltammetry and eliminated from dopamine signals using differential measurements. However, 3,4-dihydroxyphenylacetic acid (DOPAC) could contribute to our measurements: it is present in the extracellular fluid in dopamine-containing terminal regions and can be electro-oxidized. It is formed by monoamine oxidase, which operates on a minute timescale. Thus, DOPAC concentration should not change rapidly. Nonetheless, to examine pharmacologically the contribution of DOPAC to our signals we carried out some additional experiments in the presence of the monoamine oxidase inhibitor, pargyline (75 mg per kg intraperitoneal administration). The amplitudes of the signals after presentation of the audiovisual stimulus ($n = 3$) or a lever press for cocaine ($n = 1$) were not attenuated in the presence of pargyline (data not shown), confirming that DOPAC was not a contributor.

Received 28 November 2002; accepted 3 February 2003; doi:10.1038/nature01476.

- Wise, R. A. Drug-activation of brain reward pathways. *Drug Alcohol Depend.* **51**, 13–22 (1998).
- Koob, G. F. & Nestler, E. J. The neurobiology of drug addiction. *J. Neuropsychiatry Clin. Neurosci.* **9**, 482–497 (1997).
- Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
- Hyland, B. I., Reynolds, J. N. J., Hay, J., Perk, C. G. & Miller, R. Firing modes of midbrain dopamine cells in the freely moving rat. *Neuroscience* **114**, 475–492 (2002).
- Pettit, H. O. & Justice, J. B. Jr Dopamine in the nucleus accumbens during cocaine self-administration as studied by *in vivo* microdialysis. *Pharmacol. Biochem. Behav.* **34**, 899–904 (1989).
- Wise, R. A. *et al.* Fluctuations in nucleus accumbens dopamine concentration during intravenous cocaine self-administration in rats. *Psychopharmacology* **120**, 10–20 (1995).
- Di Chiara, G. Drug addiction as dopamine-dependent associative learning disorder. *Eur. J. Pharmacol.* **375**, 13–30 (1999).
- Bradberrry, C. W., Barret-Larimore, R. L., Jatlow, P. & Rubino, S. R. Impact of self-administered cocaine and cocaine cues on extracellular dopamine in mesolimbic and sensorimotor striatum in rhesus monkeys. *J. Neurosci.* **20**, 3874–3883 (2000).
- Ito, R., Dalley, J. W., Howes, S. R., Robbins, T. W. & Everitt, B. J. Dissociation in conditioned dopamine release in the nucleus accumbens core and shell in response to cocaine cues and during cocaine-seeking behavior in rats. *J. Neurosci.* **20**, 7489–7495 (2000).
- Wightman, R. M. & Robinson, D. L. Transient changes in mesolimbic dopamine and their association with ‘reward’. *J. Neurochem.* **82**, 721–735 (2002).
- Stamford, J. A. & Justice, J. B. Jr Probing brain chemistry. *Anal. Chem.* **68**, 359A–363A (1996).
- Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Breiter, H. C. *et al.* Acute effects of cocaine on human brain activity and emotion. *Neuron* **19**, 591–611 (1997).
- Childress, A. R. *et al.* Limbic activation during cue-induced cocaine craving. *Am. J. Psychiatry* **156**, 11–18 (1999).
- Wise, R. A. Neurobiology of addiction. *Curr. Opin. Neurobiol.* **6**, 243–251 (1996).
- Grace, A. A. Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience* **41**, 1–24 (1991).
- Einhorn, L. C., Johansen, P. A. & White, F. J. Electrophysiological effects of cocaine in the mesoaccumbens dopamine system: studies in the ventral tegmental area. *J. Neurosci.* **8**, 100–112 (1988).
- Lacey, M. G., Mercuri, N. B. & North, R. A. Actions of cocaine on rat dopaminergic neurones *in vitro*. *Br. J. Pharmacol.* **99**, 731–735 (1990).
- Kiyatkin, E. A. & Rebec, G. V. Dopaminergic modulation of glutamate-induced excitations of neurons in the neostriatum and nucleus accumbens of awake, unrestrained rats. *J. Neurophysiol.* **75**, 142–153 (1996).
- Pennartz, C. M., Groenewegen, H. J. & Lopes da Silva, F. H. The nucleus accumbens as a complex of functionally distinct neuronal ensembles: an integration of behavioural, electrophysiological and anatomical data. *Prog. Neurobiol.* **42**, 719–761 (1994).
- Sesack, S. R. & Pickel, V. M. In the rat medial nucleus accumbens, hippocampal and catecholaminergic terminals converge on spiny neurons and are in apposition to each other. *Brain Res.* **527**, 266–279 (1990).
- Sesack, S. R. & Pickel, V. M. Prefrontal cortical efferents in the rat synapse on unlabeled neuronal targets of catecholamine terminals in the nucleus accumbens septi and on dopamine neurons in the ventral tegmental area. *J. Comp. Neurol.* **320**, 145–160 (1992).
- Mogenson, G. J. Limbic-motor integration. *Prog. Psychobiol. Physiol. Psychol.* **12**, 117–170 (1987).
- Carelli, R. M. The nucleus accumbens and reward: neurophysiological investigations in behaving animals. *Behav. Cogn. Neurosci. Rev.* **1**, 281–296 (2002).
- Carelli, R. M., King, V. C., Hampson, R. E. & Deadwyler, S. A. Firing patterns of nucleus accumbens neurons during cocaine self-administration in rats. *Brain Res.* **626**, 14–22 (1993).
- Carelli, R. M. Activation of accumbens cell firing by stimuli associated with cocaine delivery during self-administration. *Synapse* **35**, 238–242 (2000).
- Carelli, R. M., Ijames, S. G. & Crumling, A. J. Evidence that separate neural circuits in the nucleus accumbens encode cocaine versus ‘natural’ (water and food) reward. *J. Neurosci.* **20**, 4255–4266 (2000).
- Garris, P. A., Christensen, J. R. C., Rebec, G. V. & Wightman, R. M. Real-time measurement of electrically evoked extracellular dopamine in the striatum of freely moving rats. *J. Neurochem.* **68**, 152–161 (1997).
- Marsden, C. A. *et al.* *In vivo* voltammetry—present electrodes and methods. *Neuroscience* **25**, 389–400 (1988).
- Millar, J., Stamford, J. A., Kruk, Z. L. & Wightman, R. M. Electrochemical, pharmacological and electrophysiological evidence of rapid dopamine release and removal in the rat caudate nucleus following electrical stimulation of the median forebrain bundle. *Eur. J. Pharmacol.* **109**, 341–348 (1985).

Acknowledgements We thank M. Roitman, D. Robinson, R. Gainetdinov, P. Garriss and S. Grigson for useful comments, and J. Venton, J. Peterson, C. McKinney, S. Brooks and J. Wondolowski for technical assistance. We also acknowledge the vision of R. Adams who set the foundation for this work. This work was supported by grants from the National Institute on Drug Abuse to R.M.W. and R.M.C.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M.C. (e-mail: rcarelli@unc.edu).

Control of tillering in rice

Xueyong Li[†], Qian Qian[†], Zhiming Fu^{*}, Yonghong Wang^{*}, Guosheng Xiong^{*}, Dali Zeng[‡], Xiaogun Wang^{*}, Xinfang Liu^{*}, Sheng Teng[‡], Fujimoto Hiroshi[‡], Ming Yuan[§], Da Luo^{||}, Bin Han[¶] & Jiayang Li^{*}

^{*} Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

[‡] China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310006, Zhejiang, China

[§] China Agricultural University, Beijing 100094, China

^{||} Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China

[¶] National Center for Gene Research, Chinese Academy of Sciences, Shanghai 200233, China

[†] These authors contributed equally to this work

Tillering in rice (*Oryza sativa* L.) is an important agronomic trait for grain production, and also a model system for the study of branching in monocotyledonous plants. Rice tiller is a specialized grain-bearing branch that is formed on the unelongated basal internode and grows independently of the mother stem (culm) by means of its own adventitious roots¹. Rice tillering occurs in a two-stage process: the formation of an axillary bud at each leaf axil and its subsequent outgrowth². Although the morphology and histology^{2,3} and some mutants of rice tillering⁴ have been well described, the molecular mechanism of rice tillering remains to be elucidated. Here we report the isolation and characterization of *MONOCULM 1* (*MOC1*), a gene that is important in the control of rice tillering. The *moc1* mutant plants have only a main culm without any tillers owing to a defect in the formation of tiller buds. *MOC1* encodes a putative GRAS family nuclear protein that is expressed mainly in the axillary buds and functions to initiate axillary buds and to promote their outgrowth.

To identify genes involved in the control of rice tillering, we have screened for mutants with altered tiller numbers from collections derived from spontaneous mutations or γ -ray radiation and ethyl methanesulphonate (EMS) mutagenesis. A spontaneous *monoculm 1* (*moc1*) mutant is of particular interest, because *moc1* plants nearly completely lose their tillering ability, producing only one main culm, in contrast to the multiple tillers in wild-type plants (Fig. 1). Genetic analysis with reciprocal crosses between *moc1* and wild-type plants revealed that *moc1* possesses a recessive mutation in a single nuclear locus. Allelic tests between the *moc1* mutant and five recessive tillering mutants with reduced culm number (*rcn1* to *rcn5*)⁴ indicated that *MOC1* is a previously unknown locus that is involved in the control of rice tillering.

In the seedling stage, no obvious morphological difference could be observed between *moc1* and wild-type plants. However, during the tillering stage, beginning from the fourth complete leaf formation, tillers emerged from sheaths of the subtending leaves in wild-type plants, but no tillers arose from leaf axils of *moc1* plants (Fig. 1a–d). Up to the heading stage, wild-type rice plants produced

not only primary tillers on the main culm, but also secondary ones on the primary tiller culms. In *moc1* mutants, however, no primary tillers other than a main culm could be observed, and therefore no secondary tillers were seen either (Fig. 1e, f). Similarly, *moc1* panicles also produced much fewer rachis-branches and spikelets than did wild-type plants (Fig. 1g, h). In contrast to phenotypic alterations observed in aerial organs, roots appear to be unaffected in *moc1* plants (Fig. 1a–f).

The *MOC1* locus was mapped primarily to the long arm of chromosome 6 of *O. sativa* between markers R1559 and S1437 (Fig. 2a), and was subsequently fine-mapped to a 20-kilobase (kb) region using newly developed molecular markers (Fig. 2a, b; Supplementary Table 1). Annotation of the 20-kb sequence identified an open reading frame (ORF) that encodes a protein highly homologous (44% identity) to the tomato LATERAL SUPPRESSOR (*LS*)⁵. In tomato, loss-of-function mutation in *LS* causes a branchless phenotype owing to a failure in axillary meristem initiation. This result suggests that the rice ORF with homology to the tomato *LS* is very probably the *MOC1* gene.

We therefore amplified the corresponding ORF from *moc1* and wild-type plants with polymerase chain reaction (PCR) and sequenced it. DNA sequence comparison revealed a 1.9-kb retrotransposon inserted in this ORF in the *moc1* mutant. Confirmation of the retrotransposon-interrupted ORF as *MOC1* was achieved by functional complementation. A binary plasmid carrying a 3.2-kb wild-type genomic fragment containing the entire ORF plus a 1.5-kb upstream region (pC8247), but not the one carrying a C-terminal truncation (pC8247S) (Fig. 2b), was able to rescue the monoculm phenotype of the *moc1* mutant (Fig. 1i, j). In DNA blot analysis of T₂ progeny from a self-pollinated transgenic line, the co-segregation of the transgene and the tiller phenotype is further evidence that the ORF homologous to *LS* is indeed the *MOC1* gene (data not shown). DNA blot analysis also indicated that *MOC1* is a single-copy gene in the rice genome.

The *MOC1* complementary DNA was cloned by reverse transcription (RT)–PCR using total RNA prepared from rice seedlings. Alignment of the cDNA and genomic DNA sequences revealed no introns in the *MOC1* gene, as is the case in the tomato *LS*⁵. The first in-frame ATG of the transcript (position 1) initiates an ORF encoding a protein of 441 amino-acid residues (Fig. 2c; Supplementary Fig. 1a). In *moc1*, the 1.9-kb retrotransposon is inserted at position 948, causing a premature translation stop that results in a truncated fusion protein of 338 amino-acid residues with the last 22 residues encoded by the retrotransposon sequence (Fig. 2c; Supplementary Fig. 1b).

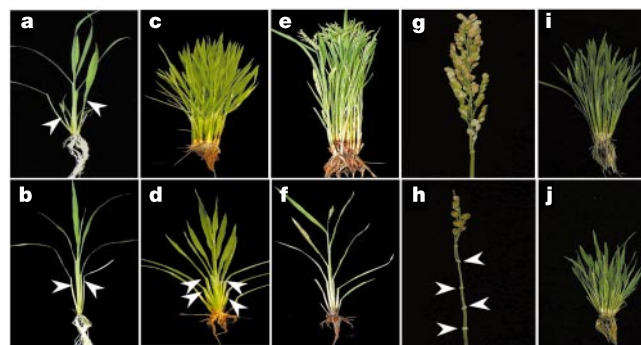


Figure 1 Phenotype and complementation of the *moc1* mutant. **a–f**, Comparison of tillering abilities between wild-type and *moc1* plants at the onset of tillering stage (**a, b**), at the peak of tillering stage (**c, d**), and at the heading stage (**e, f**). Arrows show emerging tillers in wild-type plants or empty leaf sheaths in the *moc1* mutant. **g, h**, Panicles of wild-type and *moc1* plants. Arrows show bract nodes missing rachis-branches in the *moc1* plant. **i, j**, The *moc1* plants harbouring one and three *MOC1* transgene copies.

Homology analysis demonstrated that MOC1 is a member of the plant-specific GRAS family proteins⁶, which contain several sub-families including MOC1 and LS for axillary branching⁵, SCR and SHR for root radial patterning^{7,8}, PAT1 for light signalling⁹, and GAI, RGA, SLR1, RHT1 and D8 for gibberellin-acids signalling and plant height^{10–13}. The GRAS family proteins are characterized by a conserved VHIID motif flanked by two leucine heptad repeats, a conserved C-terminal region, and an N-terminal region varying in length and sequence (Fig. 2c) that may confer specificity⁹. In addition, the MOC1 C terminus also contains an SH2-like domain¹⁴ followed by a conserved tyrosine (Fig. 2c), a potential site for phosphorylation.

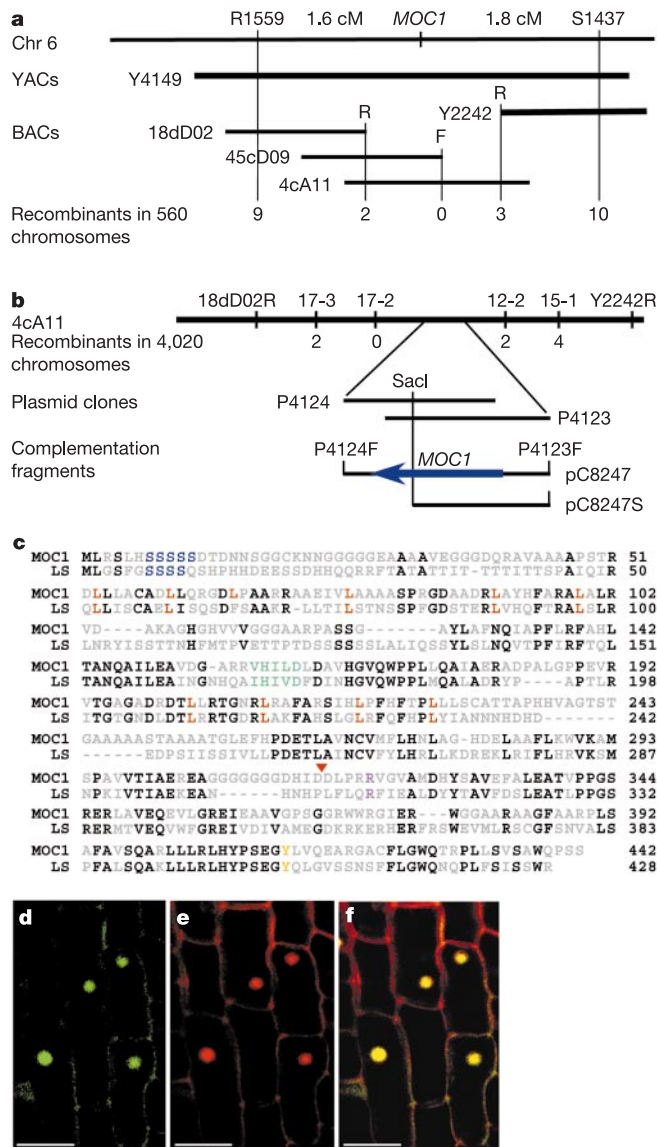


Figure 2 Molecular identification of MOC1. **a**, MOC1 was mapped within the BAC clone 4cA11. BAC, bacteria artificial chromosome; cM, centimorgans; YAC, yeast artificial chromosome. **b**, MOC1 was further localized within a 20-kb region and covered by plasmids P4123 and P4124 from the 4cA11 shotgun library. Plasmids containing the entire (pC8247) or truncated (pC8247S) MOC1 were constructed for complementation. **c**, Alignment of rice MOC1 and tomato LS. The serine homopolymeric stretch is in blue, the conserved leucine in leucine heptad repeats in red, the VHIID motif in green, the invariant arginine in SH2-like domain in purple, and the conserved tyrosine in yellow. Red triangle indicates retrotransposon insertion site. **d–f**, The MOC1–GFP fusion protein is nuclear-localized. The figure shows MOC1–GFP green fluorescence in the nucleus (**d**), the same cells stained with PI (**e**), and their merged image (**f**). Scale bars, 10 μ m.

On the basis of findings that some GRAS members such as RGA¹¹ and SLR1¹² possess nuclear localization signals (NLS), the GRAS family proteins have been proposed to function as transcription factors^{6,14}. Surprisingly, no conventional NLS domains thus far characterized¹⁵ could be predicted in MOC1 (Fig. 2c). To test whether or not MOC1 is localized in the nucleus, a GFP reporter gene encoding green fluorescent protein was fused in-frame to the last codon of MOC1 to produce a MOC1–GFP fusion protein in the transgenic rice plants. The MOC1–GFP green fluorescent signal was detected mainly in the nuclei of stable transgenic rice plants (Fig. 2d–f), indicating that MOC1 is a nuclear-localized protein and presumably functions as a transcription regulator. There may be an unidentified NLS in MOC1 or MOC1 may be imported into the nucleus through an NLS-independent mechanism.

To find out what role MOC1 plays in rice tillering, we characterized the *moc1* mutant at both anatomical and histological levels. In wild-type rice seedlings, tiller bud formation was easily recognizable along the unelongated basal internodes (Fig. 3a). In contrast, no tiller buds could be observed in the *moc1* mutant (Fig. 3b), indicating that the monocult phenotype of *moc1* is probably caused by a failure in the formation of tiller buds. In the longitudinal sections of the tip part of wild-type seedlings, tiller buds at different developmental stages could be observed (Fig. 3c, e, g, i), but not in the *moc1* mutant (Fig. 3d, f, h, j). The three tiller buds that emerged at the axils of the 6th, 5th and 4th leaves in the wild-type plants correspond approximately to the three stages of axillary bud formation proposed in *Arabidopsis*¹⁶; that is, axillary cell division

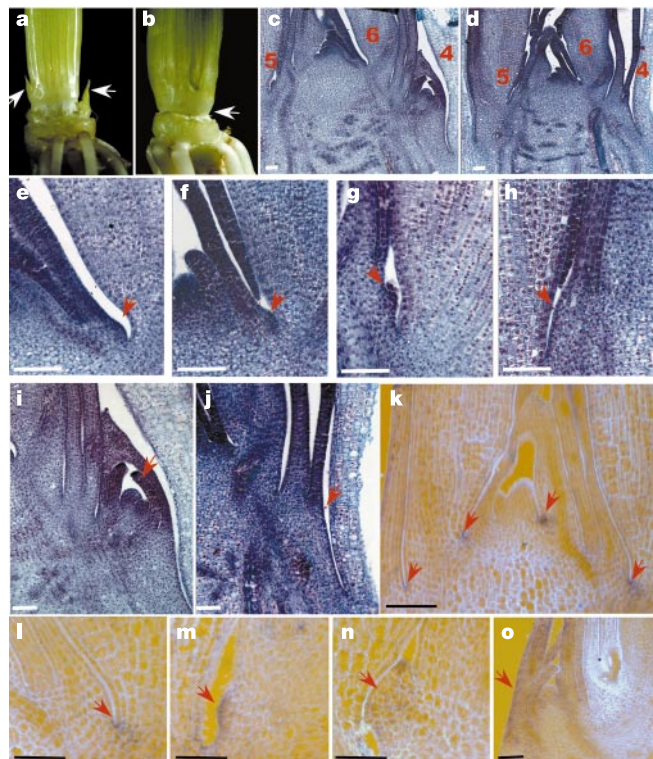


Figure 3 Microscopic studies on rice tiller bud formation. **a, b**, Tiller buds of four-leaf stage seedlings. Arrows indicate tiller buds in the wild-type and barren tillering nodes in *moc1*. **c–j**, Longitudinal sections of four-leaf stage shoot apices. **c, d**, Overall structure of wild-type and *moc1* shoot apices. Number indicates the 4th, 5th or 6th leaf. **e, g, i**, The small protuberance, axillary meristem and tiller bud formed in the 6th, 5th and 4th leaf axils in wild type. **f, h, j**, The void 6th, 5th and 4th leaf axils in *moc1*. **k–o**, RNA *in situ* hybridization of MOC1, showing specific expression in the leaf axils (**k**), in a few epidermal cells prior to any visible morphological changes (**l**), in a small protuberance before axillary meristem appearance (**m**), in axillary meristem (**n**), and in a matured tiller bud (**o**). Arrows show the MOC1 expression sites. Scale bars, 100 μ m.

(stage 1, Fig. 3e), appearance of the axillary meristems (stage 2, Fig. 3g) and differentiation of the first several axillary leaf primordia (stage 3, Fig. 3i). The defect in the formation of tiller buds in *moc1* plants could be observed as early as stage 1 (Fig. 3e, f), and no axillary meristems (Fig. 3g, h) and tiller buds (Fig. 3i, j) were formed in the *moc1* mutants. Therefore, the monocolm phenotype of *moc1* can be specifically attributed to a failure in the initiation of axillary meristems.

The requirement of *MOC1* for axillary meristem initiation and tiller bud formation was confirmed by investigating its spatial and temporal expression patterns (Fig. 3k–o). RNA *in situ* hybridization showed that *MOC1* expression could be detected in a small number of epidermal or subepidermal cells at the leaf axils, before any visible morphological changes (Fig. 3l). Subsequently, *MOC1* expression was observed in the small protuberance (Fig. 3m) and axillary meristem (Fig. 3n), and extended to the entire tiller buds including the axillary leaf primordia and young leaves (Fig. 3o). In contrast to high-level expression in the axillary meristems, *MOC1* expression was undetectable in the shoot apical meristem (SAM) (Fig. 3k). These results strongly suggest that *MOC1* has a critical role in the initiation of axillary meristems and formation of tiller buds.

In wild-type rice plants, a tiller bud is normally formed at each leaf axil, but only those formed on the unelongated basal internodes can grow out into tillers and those formed on the elongated upper internodes become arrested (Fig. 4a)¹. Secondary tillers are usually formed in wild-type plants, but higher-order tillers such as tertiary, quaternary and quinternary ones are seldom developed^{2,17} (Fig. 4c).

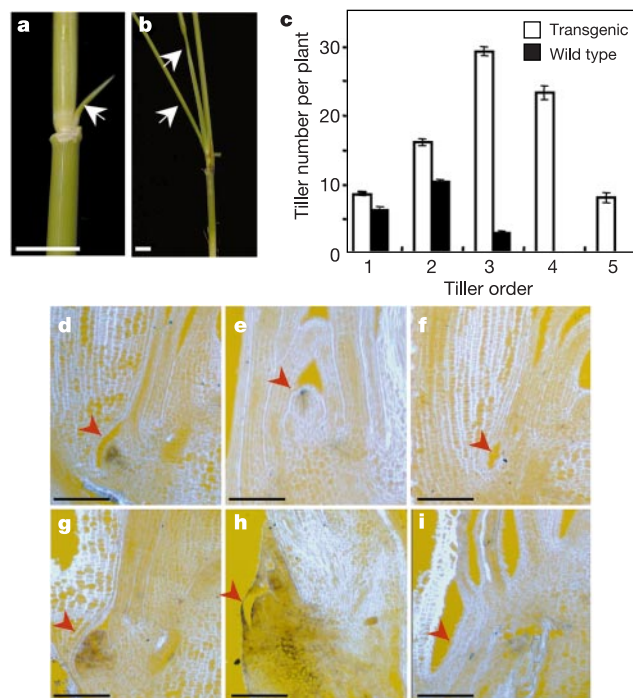


Figure 4 Promotion of *MOC1* on tiller bud outgrowth and *OSH1* and *OsTB1* expression in wild-type and *moc1* plants. **a–c**, Tiller bud outgrowth. **a**, The single arrested tiller bud (arrow) formed on the elongated upper internode of wild-type plants. **b**, Two tillers (arrows) formed on one single elongated upper internode in *MOC1* transformants. Scale bars, 1 cm. **c**, Outgrowth of higher-order tillers in *MOC1* T₂ transformants. Tiller numbers were counted at maturity stage and shown as mean $\pm$ s.e. ($n = 10$). **d–i**, *OSH1* and *OsTB1* expression in wild-type and *moc1* plants. **d–f**, *OSH1* RNA *in situ* hybridization, showing expression in axillary meristem (**d**) and in the SAM of tiller bud (**e**) in wild type, but no expression in the *moc1* leaf axil (**f**). **g–i**, *OsTB1* RNA *in situ* hybridization, showing expression in axillary meristem (**g**) and in the entire tiller bud (**h**) in wild type, but no expression in the *moc1* leaf axil (**i**). Arrows indicate the expression sites. Scale bars, 100 μ m.

However, this tillering pattern was altered in all 15 independent *MOC1* transgenic lines obtained in the complementation tests. In the *MOC1* transformants, most tiller buds formed on the elongated upper internodes (Fig. 4b), higher-order tiller buds (Fig. 4c) were able to develop into tillers and more than one tiller was formed at some leaf axils (Fig. 4b). The altered tiller phenotype was inheritable in the T₂ progeny (Supplementary Table 2).

These observations suggested that in addition to its role in the formation of tiller buds, *MOC1* also functions in promoting tiller bud outgrowth in the rice genetic background. The enhanced tillering ability of *MOC1* transformants resulted in 2–3-fold more tillers than in wild-type rice plants (Fig. 1j; Supplementary Table 2). Of the five lines we examined, the two lines harbouring three transgene copies produced many more tillers than the three other lines harbouring one copy (Supplementary Table 2), suggesting that the enhanced tillering ability was probably due to slight overexpression of the *MOC1* gene.

We have also found that the plant height of the high-tillering *MOC1* transformants was significantly reduced (Fig. 1j; Supplementary Table 2). This negative correlation between tiller number and plant height has also been observed in high-tillering dwarf mutants and cultivated rice varieties^{4,18}. The finding that *MOC1* affects both rice tillering and plant height will prompt us to investigate their molecular mechanisms further.

Tillering is a complex process in which expression of many genes must be fine-tuned. Given the importance of *MOC1* in both tiller bud formation and outgrowth, and its nuclear localization, *MOC1* is probably a key regulator in this complicated process. To identify possible downstream genes regulated by *MOC1*, we examined the expression of several important genes involved in shoot meristem or branch development. Among them, the *OSH1* gene is a key regulator of meristem initiation and maintenance¹⁹, and the *OsTB1* gene is a rice orthologue of the maize *TB1* that is expressed in axillary buds and regulates axillary bud outgrowth^{20–22}. RT-PCR analysis revealed that expression of both *OSH1* and *OsTB1* was significantly reduced in the *moc1* mutant (Supplementary Fig. 2).

This result was confirmed by *in situ* examination of *OSH1* and *OsTB1* expression patterns during tiller bud formation. In wild-type plants, besides the expression in SAM (data not shown), *OSH1* was also expressed in the axillary meristem (Fig. 4d) and the apical meristem of tiller buds (Fig. 4e). In the *moc1* mutant, however, *OSH1* expression could not be detected in the leaf axils (Fig. 4f), although the SAM expression was unaffected (data not shown). A similar observation was also made for the *OsTB1* expression in wild-type and *moc1* plants (Fig. 4g–i). These findings indicate that *OSH1* and *OsTB1* might be regulated by *MOC1* in the tillering process, and *MOC1* may act as a master regulator in the control of rice tillering.

MOC1 is the first functionally defined gene, to our knowledge, that controls tillering in rice, although several genes involved in dicotyledonous plant branching have been isolated^{5,23,24}. Given the mutant's phenotype and similarity of sequences, *MOC1* is probably the rice orthologue of tomato *LS*. However, *MOC1* has distinct functions, such as promotion of tiller bud outgrowth and a negative effect on plant height, which may reflect the fundamental difference between monocotyledonous tillering and dicotyledonous branching.

Tillering is one of the most important agronomic traits because tiller number per plant determines panicle number, a key component of grain yield¹⁸. As a key regulator of tillering, *MOC1* (and its homologues in other cereals) could make a significant contribution to future improvement of these crops. □

Methods

Plant materials

The rice *moc1* mutant was obtained from a natural occurrence in a *japonica* cultivar, H89025. A mapping population of 2,010 F₂ mutant plants was generated from the crosses between *moc1* and Minghui 63 (*indica*).

Mapping and cloning of *MOC1*

MOC1 was mapped primarily with Simple Sequence Length Polymorphism (SSLP) and Restriction Fragment Length Polymorphism (RFLP) markers²⁵, using 280 F₂ mutant plants. The *MOC1* locus was further placed within a 20-kb region between the markers 17-3 and 12-2 by using 2,010 F₂ mutant plants and newly developed molecular markers (Supplementary Table 1). The candidate gene was amplified from both the *moc1* and wild-type genomic DNA using primers MOC1F1 and MOC1R3 (5'-TCGTTGTAGTAGCTCT GGTC-3' and 5'-CTAAGTAGAGTCGAGTAGC-3'), and the PCR products were sequenced directly.

Complementation test

A 3.2-kb genomic DNA fragment containing the entire *MOC1* coding region, the 1,534-bp upstream sequence and the 316-bp downstream sequence, was inserted into the binary vector pCambia1300 to generate a transformation plasmid pC8247. A control plasmid, pC8247S, containing a 3' truncated *MOC1* gene that encodes the first 283 amino-acid residues, was also constructed. The two binary plasmids were introduced into *Agrobacterium tumefaciens* LBA4404 and the *moc1* mutant was transformed as reported previously²⁶. Fifteen independent transgenic lines were obtained for pC8247 and four for pC8247S, respectively. Each transgenic line included 10–100 sibling plants.

Subcellular localization

MOC1–GFP fusion was made by replacing the CaMV35S Ω region with a 2.9-kb DNA fragment that contained the entire *MOC1* coding region and 1.5-kb upstream sequence in the CaMV35S Ω –sGFP (S65T)–NOS-3' cassette vector²⁷. The construct was sequenced to verify the in-frame fusion and no nucleotide mutations. The *MOC1*–GFP fusion gene was subcloned into the binary vector pCambia1300 and then transformed into wild-type rice plants. The tip of the transgenic rice plant stem was sectioned longitudinally, stained for 30 min with 2 μ g ml^{−1} propidium iodide (PI) (in 30 mM 2-(N-morpholino)-ethanesulphonic acid and 100 mM mannitol, pH 5.9), and visualized with a confocal microscope (Bio-Rad MRC 1024).

Histology and *in situ* hybridization

Shoot apices of rice seedlings at the four-leaf stage were fixed with formalin–acetic acid–alcohol (FAA) fixative solution at 4 °C overnight followed by dehydration steps and then embedded in paraffin (Paraplast Plus, Sigma). The tissues were sliced into 7–10- μ m sections with a microtome (Leica RM2145), affixed to microscope slides, and stained with Safranin O and Fast Green (Fisher). Sections were observed under bright-field through a microscope (Leica DMR) and photographed using a Micro Color charge-coupled device (CCD) camera (Apogee Instruments).

RNA *in situ* hybridization was performed as described previously²⁸. The 3' ends of *MOC1*, *OsTB1* and *OSH1* were subcloned into pBluescript SK(+) vector and used as templates to generate sense and antisense RNA probes. Digoxigenin-labelled RNA probes were prepared using a DIG Northern Starter Kit (Cat. No. 2039672, Roche) according to the manufacturer's instructions. The slides were observed under bright-field and epifluorescence at the same time with a microscope (Leica DMR) and photographed using a 3CCD colour video camera (DXC-390P, Sony).

RNA extraction and analysis

Total RNA extraction was performed as described previously²⁹. The first-strand cDNA was synthesized from total RNA and used as RT–PCR templates. RT–PCR primers for *OSH1* were 5'-GAGATTGATGCACATGGTGTG-3' and 5'-ATTAGCAGCAGCAAGAGTAGC-3', for *OsTB1* were 5'-AAGTCTTCGCGCTCCAGGA-3' and 5'-AACGCCATGATCACGT CGCT-3', and for actin were 5'-TCATGAAGATCCTGACGGAG-3' and 5'-AACAGCTCC TCTGGCTTAG-3'. Rapid amplification of cDNA ends (RACE)–PCR was performed with the 5'/3' RACE Kit (Roche) according to the manufacturer's instructions.

Received 7 November 2002; accepted 19 February 2003; doi:10.1038/nature01518.

- Li, Y.-H. *Morphology and Anatomy of Grass Family Crops* 138–142 (Shanghai Science and Technology Press, Shanghai, 1979).
- Hanada, K. in *Science of the Rice Plant* Vol. 1 *Morphology* (eds Matsuo, T. & Hoshikawa, K.) 222–258 (Food and Agriculture Policy Research Center, Tokyo, 1993).
- Hanada, K. in *Science of the Rice Plant* Vol. 2 *Physiology* (eds Matsuo, T., Kumazawa, K., Ishii, R., Ishihara, K. & Hirata, H.) 61–65 (Food and Agriculture Policy Research Center, Tokyo, 1995).
- Iwata, N., Takamura, I., Wu, H.-K., Siddiqui, E. A. & Rutger, J. N. List of genes for various traits (with chromosome and main literature). *Rice Genet. Newsl.* **12**, 61–93 (1995).
- Schumacher, K., Schmitt, T., Rossberg, M., Schmitz, G. & Theres, K. The *Lateral suppressor* (*Ls*) gene of tomato encodes a new member of the VHIID protein family. *Proc. Natl Acad. Sci. USA* **96**, 290–295 (1999).
- Pysh, L. D., Wysocka-Diller, J. W., Camilleri, C., Bouchez, D. & Benfey, P. N. The GRAS gene family in *Arabidopsis*: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *Plant J.* **18**, 111–119 (1999).
- Di Laurenzio, L. *et al.* The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* **86**, 423–433 (1996).
- Helariutta, Y. *et al.* The *SHORT-ROOT* gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell* **101**, 555–567 (2000).
- Bolle, C., Koncz, C. & Chua, N. H. PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction. *Genes Dev.* **14**, 1269–1278 (2000).
- Peng, J. *et al.* The *Arabidopsis* GAI gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes Dev.* **11**, 3194–3205 (1997).
- Silverstone, A. L., Ciampaglio, C. N. & Sun, T. The *Arabidopsis* RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell* **10**, 155–169 (1998).
- Itoh, H., Ueguchi-Tanaka, M., Sato, Y., Ashikari, M. & Matsuoka, M. The gibberellin signaling pathway is regulated by the appearance and disappearance of *SLENDER RICE1* in nuclei. *Plant Cell* **14**, 57–70 (2002).

- Peng, J. *et al.* 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* **400**, 256–261 (1999).
- Richards, D. E., Peng, J. & Harberd, N. P. Plant GRAS and metazoan STATs: one family? *Bioessays* **22**, 573–577 (2000).
- Smith, H. M., Hicks, G. R. & Raikhel, N. V. Importin α from *Arabidopsis thaliana* is a nuclear import receptor that recognizes three classes of import signals. *Plant Physiol.* **114**, 411–417 (1997).
- Stirnberg, P., van De Sande, K. & Leyser, H. M. MAX1 and MAX2 control shoot lateral branching in *Arabidopsis*. *Development* **129**, 1131–1141 (2002).
- Katayama, T. Analytical studies of tillering in paddy rice. *J. Imp. Agric. Exp. Stn Jpn* **1**, 327–374 (1931).
- Yan, J.-Q., Zhu, J., He, C.-X., Benmoussa, M. & Wu, P. Quantitative trait loci analysis for the developmental behavior of tiller number in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **97**, 267–274 (1998).
- Sato, Y. *et al.* A rice homeobox gene, *OSH1*, is expressed before organ differentiation in a specific region during early embryogenesis. *Proc. Natl Acad. Sci. USA* **93**, 8117–8122 (1996).
- Doebley, J., Stec, A. & Hubbard, L. The evolution of apical dominance in maize. *Nature* **386**, 485–488 (1997).
- Lukens, L. & Doebley, J. Molecular evolution of the teosinte branched gene among maize and related grasses. *Mol. Biol. Evol.* **18**, 627–638 (2001).
- Hubbard, L., McSteen, P., Doebley, J. & Hake, S. Expression patterns and mutant phenotype of *teosinte branched 1* correlate with growth suppression in maize and teosinte. *Genetics* **162**, 1927–1935 (2002).
- Schmitz, G. *et al.* The tomato *Blind* gene encodes a MYB transcription factor that controls the formation of lateral meristems. *Proc. Natl Acad. Sci. USA* **99**, 1064–1069 (2002).
- Otsuga, D., DeGuzman, B., Prigge, M. J., Drews, G. N. & Clark, S. E. *REVOLUTA* regulates meristem initiation at lateral positions. *Plant J.* **25**, 223–236 (2001).
- Harushima, Y. *et al.* A high-density rice genetic linkage map with 2275 markers using a single F₂ population. *Genetics* **148**, 479–494 (1998).
- Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282 (1994).
- Niwa, Y., Hirano, T., Yoshimoto, K., Shimizu, M. & Kobayashi, H. Non-invasive quantitative detection and applications of non-toxic, S65T-type green fluorescent protein in living plants. *Plant J.* **18**, 455–463 (1999).
- Bobert, P. R., Coen, E. S., Murphy, G. J. & Doonan, J. H. Patterns of cell division revealed by transcriptional regulation of genes during the cell cycle in plants. *EMBO J.* **13**, 616–624 (1994).
- Hu, Y.-X., Bao, F. & Li, J. Promotive effect of brassinosteroids on cell division involves a distinct CycD3-induction pathway in *Arabidopsis*. *Plant J.* **24**, 693–701 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank J. Zuo, N.-H. Chua and X.-W. Deng for critical comments on the manuscript, B. Zhang for assistance in using the microscope, I. Takamura for providing rcn1–rcn5 mutants, the MAFF DNA Bank of Japan for providing RFLP probes and YAC clones, the Clemson University Genomic Institute for providing Nipponbare BAC library filters and clones, Y. Niwa for providing the CaMV35S Ω –sGFP (S65T)–NOS-3' construct, and M. Matsuoka for the *OSH1* cDNA clone. F.H. is a visiting scientist from the Japan International Research Center for Agricultural Sciences, Tsukuba, Japan. This work was supported by grants from the Ministry of Science and Technology of China, the National Natural Science Foundation of China and the Chinese Academy of Sciences.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.L. (e-mail: jyli@genetics.ac.cn). The GenBank accession number for *MOC1* sequences is AY242058.

The genes *orthodenticle* and *hunchback* substitute for *bicoid* in the beetle *Tribolium*

Reinhard Schröder

Interfakultäres Institut für Zellbiologie, Universität Tübingen, Abt. Genetik der Tiere, Auf der Morgenstelle 28, 72076 Tübingen, Germany

In *Drosophila*, the morphogen Bicoid organizes anterior patterning in a concentration-dependent manner by activating the transcription of target genes such as *orthodenticle* (*otd*)¹ and *hunchback* (*hb*), and by repressing the translation of *caudal*^{2,3}. Homologues of the *bicoid* gene have not been isolated in any organism apart from the higher Dipterans^{4–7}. In fact, head and thorax formation in other insects is poorly understood. To elucidate this process in a short-germband insect, I analysed

the function of the conserved genes *orthodenticle-1* (*otd-1*) and *hb* in the flour beetle *Tribolium castaneum*. Here I show that, in contrast to *Drosophila*, *Tribolium otd-1* messenger RNA is maternally inherited by the embryo. Reduction of *Tribolium otd-1* levels by RNA interference (RNAi) results in headless embryos. This shows that *otd-1* is required for anterior patterning in *Tribolium*. As in *Drosophila*, *Tribolium hb* specifies posterior gnathal and thoracic segments. The head, thorax and the anterior abdomen fail to develop in *otd-1/hb* double-RNAi embryos. This phenotype is similar to that of strong *bicoid* mutants in *Drosophila*. I propose that *otd-1* and *hb* are part of an ancestral anterior patterning system.

In *Drosophila*, *bicoid* is the central gene in the anterior patterning system. Embryos that do not express Bicoid protein fail to develop a head or thorax, and form a second telson at the anterior pole instead. Binding specificity of Bicoid to its DNA^{1,2} and RNA^{8,9} targets is mediated by a lysine (K) at position 50 of its homeo-domain (K₅₀HD). Despite its pivotal role in defining anterior pattern in *Drosophila*, *bicoid* orthologues have been isolated only from closely related higher Dipterans^{4–6}. In *Drosophila* and other higher Dipterans⁷, a *bicoid* orthologue is located in the *Hox* gene complex close to the *zerknüllt/Hox3* locus—between *proboscipedia/Hox2* and *Deformed/Hox4*. A *bicoid* orthologue is not present in this genomic interval in the beetle *Tribolium*⁷. This finding corroborates the hypothesis that *bicoid* evolved recently—probably through the divergence of a *Hox3* paralogue during evolution of the higher Dipterans^{6,10,11}—and is not part of an ancient anterior system common to more basal insects.

How is anterior patterning in the embryo organized in the absence of *bicoid*? In *Drosophila bicoid* mutants, high levels of *bicoid*-independent zygotic *hb* are able to direct the formation of a partial thorax¹². However, since *hb* cannot induce formation of head structures on its own, the conserved protein Otd, which like *Bicoid* contains a K₅₀ homeodomain, has been considered to be an ancestral head determinant¹². This hypothesis has been tested in an insect that develops as a short-germband embryo—and therefore shows a more general type of embryogenesis¹³.

First, I analysed the function of *otd-1* in the beetle *Tribolium*. In *Drosophila*, *otd* acts as a zygotic head gap gene that specifies the ocular and the antennal segments¹⁴. Of the two *otd* paralogues—*otd-1* and *otd-2*—known to exist in *Tribolium*, only *otd-1* is expressed during early blastoderm formation¹⁵. Like its orthologue in *Drosophila*, zygotic *otd-1* in *Tribolium* is expressed in an anterior position during early embryogenesis. This indicates that *otd-1*

might specify similar head segments in both species¹⁵. In contrast to *Drosophila*, *Tribolium otd-1* mRNA is provided maternally to the developing embryo. *Tribolium otd-1* mRNA is initially distributed uniformly in the egg, and is present at the beginning of blastoderm formation (Fig. 1a–d). During blastoderm stages, *otd-1* mRNA and protein recede first from the posterior (Fig. 1i–l) and later also from the anterior pole, resulting in a belt-like domain in the prospective head region (Fig. 1e). Dorsal *otd-1* mRNA then begins to disappear (Fig. 1f), leading to a ventrally located head stripe.

In *Tribolium*, RNAi has been shown to phenocopy gene disruption¹⁶. To investigate the function of *otd-1* during early pattern formation in *Tribolium*, I injected double-stranded (ds) *otd-1* RNA into eggs (*otd-1*^{eRNAi}) or female pupae (*otd-1*^{pRNAi}) (ref. 17). By applying pRNAi, I specifically aimed to knock down maternally supplied *otd-1* RNA and to avoid injection artefacts. Of the larval cuticles analysed, 91% of *otd-1*^{eRNAi} and 100% of *otd-1*^{pRNAi} developed defects in anterior patterning (Fig. 2a and Table 1). In the most extreme cases, the embryos lacked all head structures (Fig. 3e), indicating that *Tribolium otd-1* is required for head formation. Depending on severity, these cuticles can be grouped into classes I–VI—from weakly to strongly affected larvae (Figs 2c–h and 3c, d, e and Table 1). The range of cuticle defects is reminiscent of weak and intermediate *bicoid* mutants in *Drosophila*¹⁸. Moreover, head defects are detectable before cuticle formation in embryos, as determined by an analysis of Engrailed expression (see Fig. 4b and Supplementary Information). This shows that *Tribolium otd-1* is required for anterior patterning early during embryogenesis.

A significantly greater number of the more strongly affected class III and IV embryos was observed after pRNAi, compared with eRNAi (Fig. 2a and Table 1). The pRNAi depletes most of the maternal and zygotic *otd-1* mRNA, whereas eRNAi seems to deplete only zygotic *otd-1* mRNA. So, *otd-1* seems to be required to define the presumptive head region in *Tribolium*. Furthermore, maternal *otd-1* displays functional characteristics of a coordinate gene, as *bicoid* does in *Drosophila*, rather than of a canonical head gap gene. In *Tribolium* and *Drosophila*, zygotic *otd* seems to act as a head gap gene.

otd-1 is also involved in the formation of the anterior-most region of the egg—the extraembryonic serosa. In *otd-1*^{pRNAi} embryos, the nuclei of the serosa are irregularly arranged or, in more severe cases, reduced in number (see Supplementary Information). *hb*, which is strongly expressed in the serosa of the wild-type embryos¹⁹, is also still detectable in the nuclei of the serosa in *otd-1*^{pRNAi} embryos. *otd-1* is therefore not required to activate *hb* expression in the

Table 1 Phenotypes obtained in the *otd-1*, *hb* and *otd-1/hb* RNAi experiments

<i>otd-1</i> RNAi cuticles								
	WT	Weak to strong head defects				Headless ϕ		<i>Dll</i> ϕ
		I	II	III	IV	V	VI	
Experiment	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
<i>otd-1</i> eRNAi (n = 100)	9 (9)	1 (1)	45 (45)	25 (25)	10 (10)	6 (6)	4 (4)	-
Control: H ₂ O injection into eggs (n = 74)	96 (71)	3 (2*)	-	1 (1)	-	-	-	-
<i>otd-1</i> pRNAi (n = 111)	-	-	8 (9)	52 (58)	35 (39)	3 (3)	2 (2)	-
Control: H ₂ O injection into pupae (n = 32)	100 (32)	-	-	-	-	-	-	-
Control: <i>Dll</i> pRNAi (n = 43)	-	-	-	-	-	-	-	100 (43)
<i>hb</i> pRNAi cuticles								
Experiment	WT % (n)	Weak ϕ % (n)	Gap ϕ % (n)	8 AS % (n)	7 AS % (n)	6 AS % (n)		
<i>hb</i> pRNAi (n = 168)	11 (19)	39 (65)	50 (84)	67 (56)	21 (18)	12 (10)		
<i>otd-1/hb</i> pRNAi embryos stained for Engrailed expression								
Experiment	WT % (n)	<i>hb</i> ϕ % (n)	Headless ϕ % (n)	6 AS % (n)	5 AS % (n)	4 AS % (n)	3 AS % (n)	2 AS % (n)
<i>otd-1/hb</i> pRNAi (n = 61)	33 (20)	1.5 (1)	65.5 (40)	2.5 (1)	12.5 (5)	35 (14)	30 (12)	20 (8)

The *otd-1*^{pRNAi} phenotype is variable. The cuticles obtained are grouped into classes I–VI (also shown in Fig. 2c–h). Class V and VI show the headless phenotype (headless ϕ). In the water-injected control embryos, two (*) without eyes but with intact antennae were also grouped as class I cuticles. For the *Distal-less* phenotype (*Dll* ϕ), see also ref. 17. The *hb*^{pRNAi} gap phenotype (see Fig. 3f) was observed in 50% of the analysed cuticles. Weaker phenotypes (weak ϕ) have also been obtained and will be described elsewhere. The number of abdominal segments (AS) varies in embryos that display the *hb*^{pRNAi} gap phenotype. Presumably, the abdominal segments A8 and A7 are missing in embryos with only six abdominal segments, since *hb* is also expressed in the primordia of these segments¹⁹. Cuticles with fewer abdominal segments were not found. *otd-1/hb*^{pRNAi} embryos were fixed and immunohistochemically stained for the expression of Engrailed. 65.5% of the analysed cuticles have the headless phenotype (headless ϕ). The number of abdominal segments varies in these headless embryos from two to six.

serosa. Nevertheless, *otd-1* could assist other factors in activating *hb* expression, such as *zerknüllt*, which is also expressed in this tissue²⁰.

The fact that the *otd-1*^{RNAi} phenotype only incompletely mimics the *Drosophila bicoid* mutant phenotype implies that *otd-1* is only a partial functional equivalent of *bicoid* in *Tribolium*. I disrupted *hb* expression by pRNAi to determine whether this would result in a phenotype overlapping with that of *otd-1*^{RNAi}. Disruption of *hb* mRNA resulted in a lack of the maxillary, the labial and all three thoracic segments in 50% of the analysed embryos (see Fig. 3f,

Table 1 and Supplementary Information), whereas the anterior-most head segments all developed normally. *Drosophila* embryos that lack maternally derived and zygotic *hb* have the same phenotype²¹ (E. A. Wimmer and C. Desplan, personal communication). This indicates that, in *Drosophila* and *Tribolium*, *hb* acts as a canonical head gap gene. In the wasp *Nasonia*, the *hb* mutant phenotype is more severe, in that the complete head (except for the anterior-most labrum) and the thorax are missing²². So, the function of *hb* as a gap gene has been conserved throughout evolution.

The head defects observed in the *otd-1*^{RNAi} experiments indicate that *otd-1* and *hb* are both involved in regulating the development of gnathal segments. To evaluate the extent of their overlapping functions, embryos were generated in which both *otd-1* and *hb* expression were disrupted by pRNAi. These embryos were called

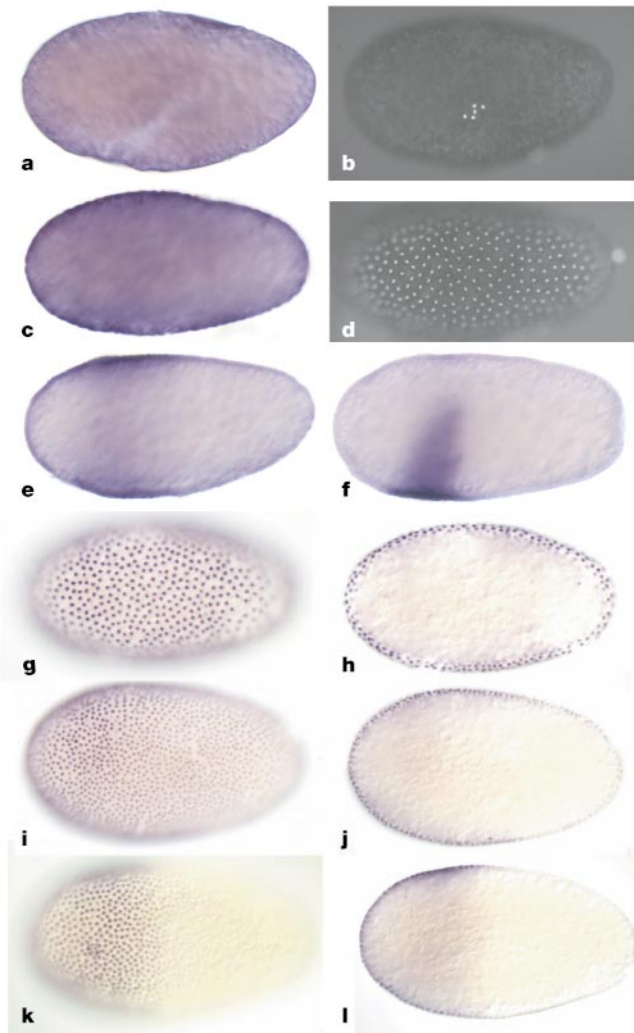


Figure 1 *otd-1* expression during blastoderm formation in *Tribolium*. **a–f**, *otd-1* mRNA expression (optical sections; anterior to the left). **a**, Uniform distribution of maternal *otd-1* mRNA in a freshly laid egg. **b**, The same egg as in **a**, but with 4',6'-diamidino-2-phenylindole hydrochloride (DAPI) staining. **c**, *otd-1* mRNA remains evenly distributed during the early stages of blastoderm formation. **d**, The same egg as in **c**, but with DAPI staining. **e**, At a later blastoderm stage, *otd-1* mRNA retracts first from the posterior and then from the anterior egg pole. Presumably, at around this stage, maternal expression starts to overlap with increasing levels of zygotic expression. **f**, Ventral head-specific *otd-1* is expressed in a wedge-like domain only after the beginning of serosa differentiation (anterior to the *otd-1* stripe). At this stage, *otd-1* is probably expressed solely from zygotic mRNA. **g–l**, Otd-1 protein expression (**g**, **i**, **k**, surface view; **h**, **j**, **l**, optical section). **g**, **h**, Uniform expression of Otd-1 protein at the beginning of blastoderm formation. **i**, **j**, Otd-1 protein recedes from the posterior pole, leading to a transient gradient that spans the posterior half of the embryo. **k**, **l**, *otd-1* mRNA (as previously described¹⁵) and protein are expressed as an anterior cap. A short-range gradient is seen in the centre of the egg at the border between expressing and non-expressing cells.

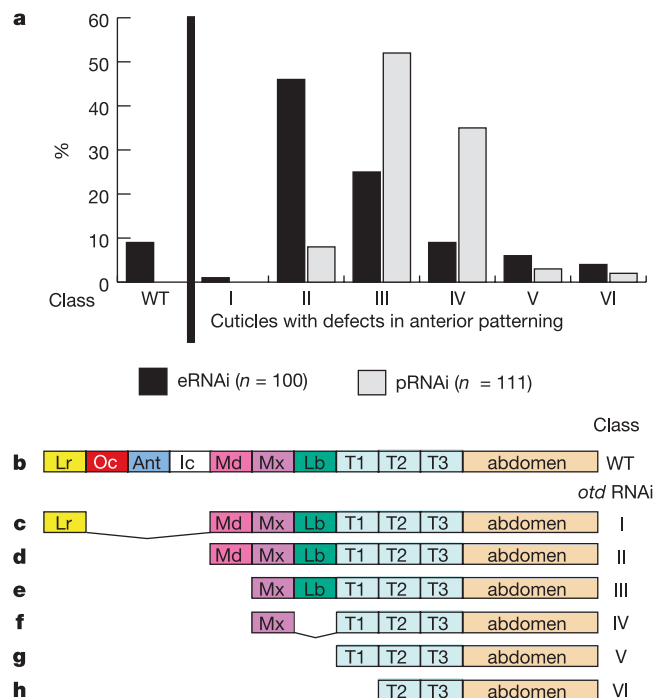


Figure 2 Effect of *otd-1*^{RNAi} on embryonic development. **a**, Analysis of the cuticle phenotype of *otd-1*^{eRNAi} and *otd-1*^{pRNAi} larvae. The same batch of dsRNA was used in both experimental series. WT, wild type. **b**, Organization of head and trunk: Lr, labrum; Oc, ocular segment with the eyes; Ant, antennae; Md, mandibles; Mx, maxillae; Lb, labium; T1–3, thoracic segments 1 to 3; abdomen. The intercalary (Ic) segment between the antennal and mandibular segment is not represented in the larval cuticle, but was added for completeness. **c–h**, Schematic representation of *otd-1*^{RNAi} phenocopies. **c**, *otd-1*^{RNAi} class I: gap phenotype. In *Drosophila*, this phenotype is seen in strong *otd* mutants. According to this phenotype, *Drosophila otd* was classified as a head gap gene¹⁴. Class I embryos frequently occur in other *otd-1*^{RNAi} series using different *otd-1* dsRNA preparations. **d**, *otd-1*^{RNAi} class II: pregnathal head missing. As in class I, it cannot be judged whether the intercalary segment is present. **e**, *otd-1*^{RNAi} class III: anterior-most gnathal segment (mandible) missing. **f**, *otd-1*^{RNAi} class IV: maxillary segment fused to thorax. Of the gnathum, the maxilla is the most stable segment with respect to the reduction of *otd-1* function. This indicates that parallel to *otd-1*, other genes are involved in specifying this segment. As judged from the analysis of *otd-1*^{RNAi} embryos stained for Engrailed expression (see Supplementary Information), the appendage seen is not a transformation of the labium. Rather, the labium is lost, leaving the maxilla in front of the thorax. **g**, *otd-1*^{RNAi} class V: headless phenotype; thorax and abdomen complete. **h**, *otd-1*^{RNAi} class VI: headless phenotype; thorax 1 missing; thoracic segment 2 and 3 formed normally; abdomen complete. Defects in thoracic segment 1 are probably not directly related to missing *otd-1* function, but rather the indirect consequence of defects in anterior patterning on more posterior patterning.

otd-1/hb^{PRNAi} double phenocopies. Forty out of 61 (65.5%) of these embryos developed a headless phenotype (Table 1), indicating that *otd-1* and *hb* function together to regulate head development.

Analysis of the Engrailed expression pattern in the *otd-1/hb*^{PRNAi} embryos revealed that only two to six abdominal segments of normal polarity develop in *otd-1/hb*^{PRNAi} embryos (see Figs 3g and 4d, Table 1 and Supplementary Information). This shows that *otd-1* and *hb* not only direct the development of head and thorax, but also act synergistically during the segmentation process of the anterior abdomen—possibly by regulating posterior gap and/or homeotic genes. A synergistic relationship between *bicoid* and *hb* has been described previously in *Drosophila*¹⁰, and might represent an evolutionary principle to confer robustness to the segmentation process. In *Tribolium*, *Otd-1* and *Hb* might substitute for the function of *Bicoid* as a transcriptional activator in *Drosophila*.

In *Drosophila*, *caudal* is translationally repressed in the anterior half of the egg by Bicoid²³. Since no ectopic posterior structures form at the anterior pole of *otd-1/hb*^{PRNAi} embryos, *Otd-1* might not serve as a translational repressor in *Tribolium*, as Bicoid does in

Drosophila. In *Drosophila*, the RNA-binding activity of Bicoid—and therefore its ability to inhibit *caudal* translation—requires the presence of an arginine at residue 54 of its homeodomain 23. This amino acid is replaced by an alanine in the homologous position of *Tribolium Otd-1*, so its RNA- but not DNA-binding specificity might be lost. A repressor of *Tribolium caudal* is therefore likely to exist in the beetle, and serves as a further component of the anterior development system. Such a repressor has yet to be identified, but different mechanisms of *caudal* regulation have been described in other systems^{24,25}. So, during Dipteran evolution, the function of Bicoid seems to have replaced the function of maternally inherited *otd-1*, and that of a repressor of *caudal*.

The expression of *otd-1* (Fig. 1) might be controlled early in embryogenesis by repression, and during later stages by autoactivation. Maternal *otd-1* mRNA translation could be repressed at the posterior pole, in a manner similar to that of maternal *hb* in *Drosophila*. In *Drosophila*²⁶, and similarly in the grasshopper *Schistocerca*²⁷, *hb* expression is repressed by the Pumilio/Nanos complex. A predicted Nanos-response element (NRE), identified by the sequence TgGTTGTattatAATTGTAaggTA (position 1429–1451, capital letters indicating identity to the NRE of *hb* in *Drosophila* and *Musca*²⁸), is present in the 3' untranslated region of *Tribolium otd-1*. Although no orthologue of *pumilio* or *nanos* has been isolated from *Tribolium*, their function in *Tribolium* has already been implicated by the downregulation of maternal *hb* at the posterior pole of the blastoderm embryo¹⁹. Maternal *otd-1* could serve to activate the expression of zygotic *otd-1* only in the

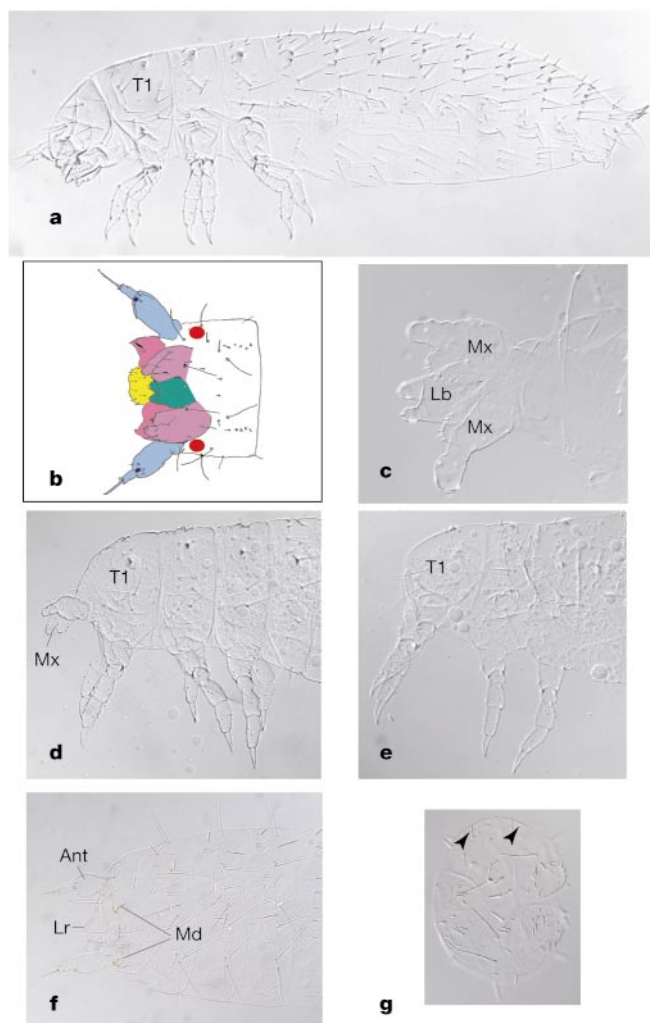


Figure 3 Cuticle preparations of wild-type and RNAi embryos. **a**, Wild-type cuticle of a first-instar larva. The thorax bears three pairs of larval legs. **b**, Cuticle of a wild-type head (camera lucida drawing of a first-instar larva). The appendages of each head segment are colour-coded, according to Fig. 2b. The eyes, which can be recognized before embedding, were added. **c**, *otd-1*^{RNAi} class III cuticle: compare to Fig. 4b. **d**, *otd-1*^{RNAi} class IV cuticle. **e**, *otd-1*^{RNAi} class V cuticle: headless phenotype. **f**, *hb*^{RNAi} cuticle. **g**, *otd-1/hb*^{RNAi} cuticle. Two segments are represented by two tracheal openings (arrowheads).

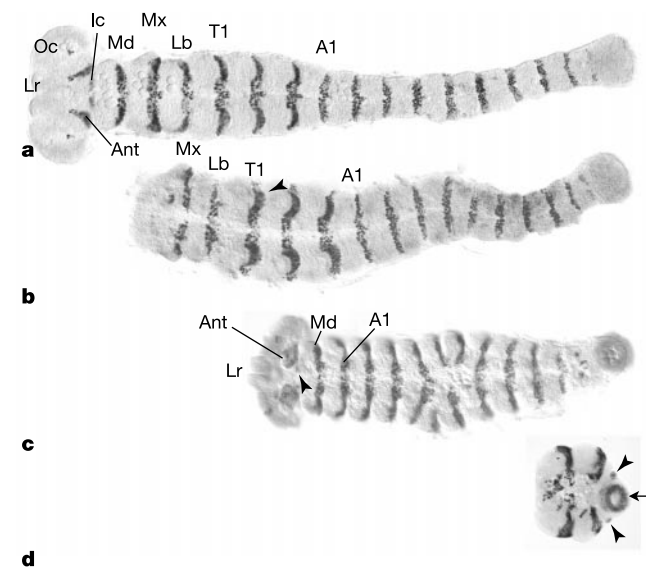


Figure 4 Engrailed staining of wild-type and RNAi embryos. *Engrailed* expression is seen in the posterior segmental compartment, and corresponds to the following segments: Lr, labrum; Ant, antenna; Oc, ocular segment; Ic, intercalary segment; Md, mandible; Mx, maxilla; Lb, labium; T1, thoracic segment 1; A1, abdominal segment 1. **a**, Wild-type embryo. The appendages of the gnathal segments and the thorax start to bulge out (anterior points to the left). **b**, *otd-1*^{PRNAi} embryo (class III). The gnathal and thoracic appendages are elongated (arrowhead points to the growing leg in T1). The maxillary segment can be identified by its drumstick-like appearance. Engrailed-positive spots, anterior to the maxillary *Engrailed* stripe, mark parts of a segment that is not properly organized. **c**, *hb*^{PRNAi} embryo. Identity of the palps, anterior to abdominal segment 1 (A1), as the mandibles is inferred from an analysis of the cuticle (Fig. 3f). Segmentation defects in the central abdomen are frequently seen. The intercalary segment is represented by a few Engrailed-positive cells posterior to the antennae (arrowhead). **d**, *otd-1/hb*^{PRNAi} embryo: extreme phenotype with two abdominal *Engrailed* stripes of normal polarity. Anlagen for the hindgut (arrow) and the malpighian tubules (arrowheads) develop only at the posterior end.

embryonic region, at the border of the serosa. The zygotic expression could then be maintained by an autoregulatory loop. To test this possibility, *otd-1*-binding sites in its regulatory region will have to be identified, and the effects of disrupting this site will have to be tested *in vitro* and *in vivo*.

As previously proposed^{11,29}, *bicoid* might have evolved from a duplicated *zerknüllt*-like gene, which was converted by mutation into a gene that codes for a K₅₀HD-containing protein—thus adopting the same DNA-binding specificity as Otd-1. During evolution of the higher Diptera, *otd* expression came under the regulation of the maternal gene *bicoid*, so that maternal Otd-1 function was not longer required. Thus, *otd* was restricted to become a gap gene in insects derived from this lineage. □

Methods

GenBank accession number AJ223627 (*Tribolium castaneum* mRNA for Orthodenticle-1 protein) was used in this study.

Antibody production and staining

A polyclonal antibody was produced by immunizing rabbits with bacterially expressed *Tribolium* Orthodenticle-1 protein (Eurogentec). The coding region of the protein used in immunization corresponds to position 760–1474 of the mRNA¹⁵, which was cloned into the *PvuII* site of the PRSET A vector (Invitrogen). Immunohistochemistry for Otd-1 and Engrailed (using the 4D9 antibody) expression, was carried out as described previously¹⁵. *In situ* hybridizations were performed according to ref. 19; the embryos shown in Fig. 1a–f are from the same immunohistochemical reaction and were processed identically.

RNA interference

dsRNA was produced as described³⁰ using the Maxi-Script Kit (Ambion). The *otd-1* dsRNA was produced from a complementary-DNA deletion clone (ΔPst), representing the 3' part of the RNA (position 1215–1991)¹⁵, which does not contain the homeobox. The cDNA clone that contains the complete open reading frame, and another deletion construct, ΔXho, which represents the 5' part of the RNA (position 1–918)¹⁵, included the homeobox and were used as controls (not shown). *Tribolium* *hb* dsRNA was produced from a 2.4-kilobase cDNA clone¹⁹. Eggs were collected at intervals of 2 h and were microinjected using a standard injection apparatus. At the time of injection, embryos were at an early stage of nuclear division. For parental RNAi, female pupae were immobilized with heptane glue in a Petri dish lid, and injected by free hand under a stereomicroscope, using a glass capillary held by a needle holder connected to an air-filled syringe. dsRNA (500 ng μl⁻¹) was injected to produce *otd-1*^{RNAi} or *hb*^{RNAi} embryos, and equal amounts of both dsRNA preparations (1 μg μl⁻¹) were injected to produce *otd-1/hb*^{RNAi} embryos. *Distal-less* dsRNA (900 base pairs) was injected at a concentration of 1 μg μl⁻¹ as a control. Only the appendage-specific phenotype¹⁷ was obtained in all of the analysed cuticles (*n* = 43), irrespective of the higher number of molecules injected.

Received 2 September 2002; accepted 25 February 2003; doi:10.1038/nature01536.

- Gao, Q. & Finkelstein, R. Targeting gene expression to the head: the *Drosophila orthodenticle* gene is a direct target of the Bicoid morphogen. *Development* **125**, 4185–4193 (1998).
- St Johnston, D. & Nüsslein-Volhard, C. The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201–220 (1992).
- Lall, S. & Patel, N. H. Conservation and divergence in molecular mechanisms of axis formation. *Annu. Rev. Genet.* **35**, 407–437 (2001).
- Sommer, R. & Tautz, D. Segmentation gene expression in the housefly *Musca domestica*. *Development* **113**, 419–430 (1991).
- Schröder, R. & Sander, K. A comparison of transplantable *bicoid* activity and partial *bicoid* homeobox sequences in several *Drosophila* and blowfly species (Calliphoridae). *Wilhelm Roux Arch. Dev. Biol.*, **203**, 34–43 (1993).
- Staubert, M., Jäckle, H. & Schmidt-Ott, U. The anterior determinant *bicoid* of *Drosophila* is a derived Hox class 3 gene. *Proc. Natl Acad. Sci. USA* **96**, 3786–3789 (1999).
- Brown, S. J. *et al.* A strategy for mapping *bicoid* on the phylogenetic tree. *Curr. Biol.* **11**, R43–R44 (2001).
- Dubnau, J. & Struhl, G. RNA recognition and translational regulation by a homeodomain protein. *Nature* **379**, 694–699 (1996).
- Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. J. & Jäckle, H. RNA binding and translational suppression by *bicoid*. *Nature* **379**, 746–749 (1996).
- Simpson-Brose, M., Treisman, J. & Desplan, C. Synergy between the *hunchback* and *bicoid* morphogens is required for anterior patterning in *Drosophila*. *Cell* **78**, 855–865 (1994).
- Dearden, P. & Akam, M. Axial patterning in insects. *Curr. Biol.* **9**, R591–R594 (1999).
- Wimmer, E., Carleton, A., Harjes, P., Turner, T. & Desplan, C. *bicoid*-independent formation of thoracic segments in *Drosophila*. *Science* **287**, 2476–2479 (2000).
- Tautz, D., Friedrich, M. & Schröder, R. Insect embryogenesis—what is ancestral and what is derived? *Development*, (Suppl.), 193–199 (1994).
- Cohen, S. & Jürgens, G. *Drosophila* headlines. *Trends Genet.* **7**, 267–272 (1991).
- Li, Y. *et al.* Two *orthodenticle*-related genes in the short-germ beetle *Tribolium castaneum*. *Dev. Genes* **206**, 35–45 (1996).
- Brown, S. J., Mahaffey, J., Lorenzen, M., Denell, R. & Mahaffey, J. Using RNAi to investigate orthologous homeotic gene function during development of distantly related insects. *Evol. Dev.* **1**, 11–15 (1999).
- Bucher, G., Scholten, J. & Klingler, M. Parental RNAi in *Tribolium* (Coleoptera). *Curr. Biol.* **12**, R85–R86 (2002).

- Frohnhofer, H. G. & Nüsslein-Volhard, C. Organization of anterior pattern in the *Drosophila* embryo by the maternal gene *bicoid*. *Nature* **324**, 120–125 (1986).
- Wolff, C., Sommer, R., Schröder, R., Glaser, G. & Tautz, D. Conserved and divergent expression aspects of the *Drosophila* segmentation gene *hunchback* in the short germ band embryo of the flour beetle *Tribolium*. *Development* **121**, 4227–4236 (1995).
- Falciani, F. *et al.* Class 3 Hox genes in insects and the origin of *zen*. *Proc. Natl Acad. Sci. USA* **93**, 8479–8484 (1996).
- Lehmann, R. & Nüsslein-Volhard, C. *hunchback*, a gene required for segmentation of an anterior and posterior region of the *Drosophila* embryo. *Dev. Biol.* **119**, 402–417 (1987).
- Pultz, M., Pitt, J. & Alto, N. Extensive zygotic control of the anteroposterior axis in the wasp *Nasomia vitripennis*. *Development* **126**, 701–710 (1999).
- Niessing, D. *et al.* Homeodomain position 54 specifies transcriptional versus translational control by Bicoid. *Mol. Cell* **5**, 395–401 (2000).
- Draper, B. W., Mello, C. C., Bowerman, B., Hardin, J. & Priess, J. R. MEX-3 is a KH domain protein that regulates blastomere identity in early *C. elegans* embryos. *Cell* **87**, 205–216 (1996).
- Isaacs, H., Andreazzoli, M. & Slack, J. Anteroposterior patterning by mutual repression of *orthodenticle* and *caudal*-type transcription factors. *Evol. Dev.* **1**, 143–152 (1999).
- Gamberi, C., Peterson, D. S., He, L. & Gottlieb, E. An anterior function for the *Drosophila* posterior determinant Pumilio. *Development* **129**, 2699–2710 (2002).
- Lall, S., Ludwig, M. Z. & Patel, N. H. *nanos* plays a conserved role in axial patterning outside of the Diptera. *Curr. Biol.* **13**, 224–229 (2003).
- Patel, N. *et al.* Grasshopper *hunchback* expression reveals conserved and novel aspects of axis formation and segmentation. *Development* **128**, 3459–3472 (2001).
- Staubert, M., Prell, A. & Schmidt-Ott, U. A single *Hox3* gene with composite *bicoid* and *zerknüllt* expression characteristics in non-Cyclorhaphan flies. *Proc. Natl Acad. Sci. USA* **99**, 274–279 (2002).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements I thank T. Mader for excellent technical assistance, A. Beermann, H. Dove, F. Maderspacher, R. Reuter and C. Wolff for critically reading drafts of the manuscript, E. A. Wimmer for discussions and for pointing out the existence of an NRE site in the *otd-1* sequence, and the Deutsche Forschungsgemeinschaft for financial support.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for material should be addressed to the author (e-mail: reinhard.schroeder@uni-tuebingen.de).

The BMP antagonist noggin regulates cranial suture fusion

Stephen M. Warren*, Lisa J. Brunet†, Richard M. Harland‡, Aris N. Economides‡ & Michael T. Longaker*

* Department of Surgery, Stanford University School of Medicine, Stanford, California 94305-5148, USA

† Department of Molecular and Cell Biology, Division of Biochemistry and Molecular Biology, University of California, Berkeley, California 94720-3204, USA

‡ Regeneron Pharmaceuticals, Inc., Tarrytown, New York 10591-6707, USA

During skull development, the cranial connective tissue framework undergoes intramembranous ossification to form skull bones (calvaria). As the calvarial bones advance to envelop the brain, fibrous sutures form between the calvarial plates¹. Expansion of the brain is coupled with calvarial growth through a series of tissue interactions within the cranial suture complex². Craniosynostosis, or premature cranial suture fusion, results in an abnormal skull shape, blindness and mental retardation³. Recent studies have demonstrated that gain-of-function mutations in fibroblast growth factor receptors (*fgfr*) are associated with syndromic forms of craniosynostosis^{4,5}. Noggin, an antagonist of bone morphogenetic proteins (BMPs), is required for embryonic neural tube, somites and skeleton patterning^{6–8}. Here we show that *noggin* is expressed postnatally in the suture mesenchyme of patent, but not fusing, cranial sutures, and that *noggin* expression is suppressed by FGF2 and syndromic *fgfr*

signalling. Since *noggin* misexpression prevents cranial suture fusion *in vitro* and *in vivo*, we suggest that syndromic *fgfr*-mediated craniosynostoses may be the result of inappropriate downregulation of *noggin* expression.

In the mouse, the posterior frontal cranial suture fuses in the first 45 days of life, whereas all other sutures, including the sagittal and coronal, remain patent⁹. The dura mater regulates cranial suture fusion through multiple pathways⁹, including signalling by transforming growth factor- β 1 (refs 9, 10) and fibroblast growth factor-2 (FGF2)¹¹ from the posterior frontal dura mater. To address whether posterior frontal suture fusion is indeed an active process, when compared with sagittal or coronal suture patency, we examined other factors known to promote bone formation, such as BMPs^{12,13}.

Surprisingly, we found abundant BMPs in both fusing and patent cranial sutures on gene chip analysis and in quantitative real-time polymerase chain reaction with reverse transcription (RT-PCR) (data not shown). Immunohistochemistry localized BMP4 protein to the suture mesenchyme and osteogenic fronts of the posterior frontal suture 1 week before the onset of fusion, as well as to patent sutures (Fig. 1a and b). The presence of BMP4 in fusing and non-fusing sutures suggested that there might be suture-specific regulation of BMP activity by secreted BMP antagonists^{6,7,14–16}.

We therefore screened for BMP antagonist messenger RNAs before (day 15), during (days 35 and 42), and after (day 50) the period of predicted suture fusion using *in situ* hybridization. We did not detect *dan/cerberus* family antagonists (*gremlin*, *dcrl* (GenBank accession number AA289245), *prdc*¹⁷, *sost*¹⁸, *dcrl* (GenBank accession number BC021458) or *cerberus*) in either patent or fusing sutures; however, we detected *noggin* in patent sutures (data not shown). To facilitate the temporal and spatial analysis of *noggin* expression, we used in-frame *lacZ/noggin* transgenic mice⁶. Stage-specific *lacZ* staining revealed *noggin* expression in the patent sagittal and coronal sutures before, during, and after the period of

predicted suture fusion (Fig. 1c). In marked contrast, there was almost no *noggin* expression in the fusing posterior frontal suture complex as early as day 15 (Fig. 1c).

Because BMP4 is present in both fusing and patent sutures, and osteoblasts line the osteogenic fronts of these sutures, we examined the effects of BMP4 on *Noggin* expression in osteoblasts. Primary calvarial osteoblasts treated with BMP4 expressed *Noggin* protein in a dose-dependent manner (Fig. 2a). If BMP4 induces *Noggin* expression, how can the posterior frontal suture fuse? Significantly, only the posterior frontal dura mater expresses FGF2 mRNA and protein *in vivo*⁹ and produces high levels of FGF2 in culture (Fig. 2b). FGF2 might therefore regulate BMP4-induced *Noggin* expression in calvarial osteoblasts¹⁹. Interestingly, FGF2 disrupts *Noggin* induction in a dose-dependent fashion (Fig. 2c). This suggests that environments with a low FGF2 concentration (that is, sagittal or

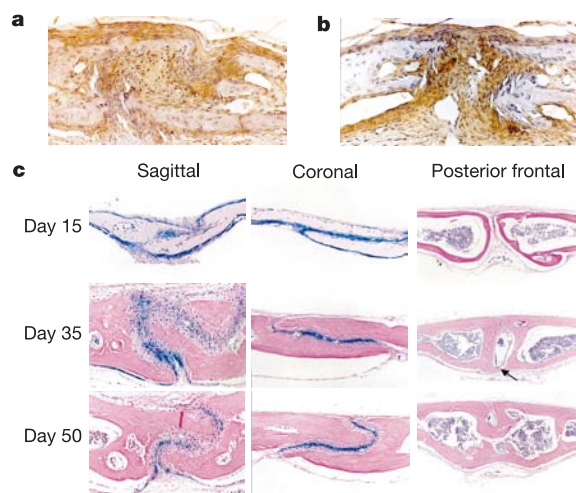


Figure 1 BMP4 and *noggin* expression in patent and fusing sutures. **a**, Photomicrograph of BMP4 immunolocalization in 18-day-old patent sagittal suture. Note staining in the suture mesenchyme and cells lining the osteogenic fronts. Controls, from which primary antibodies were omitted, yielded no detectable staining. **b**, Photomicrograph of BMP4 immunolocalization in 18-day-old posterior frontal suture, 1 week before the onset of suture fusion. Note the staining in the suture mesenchyme and cells lining the osteogenic fronts. Controls, from which primary antibodies were omitted, yielded no detectable staining. **c**, Ontogeny of *noggin* expression in the patent sagittal and coronal sutures compared with the fusing posterior frontal suture. Sections from sutures in *lacZ/noggin* transgenic mice were stained on the indicated days (day 15 is before onset of posterior frontal suture fusion; day 35 is during posterior frontal suture fusion; day 50 is after posterior frontal suture fusion). The posterior frontal suture fuses in an endocranial to ectocranial direction. Arrow indicates initial endocranial bridging (posterior frontal day 35).

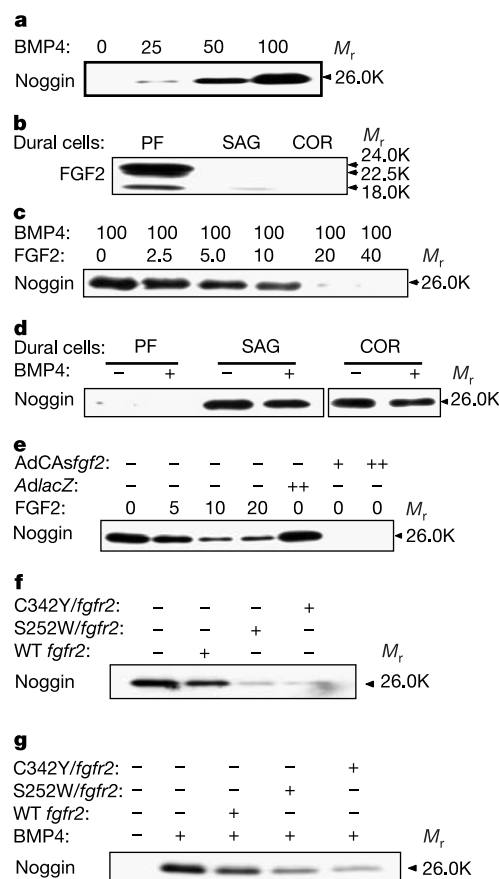


Figure 2 BMP4-induced *Noggin* expression in primary osteoblasts and dural cells is suppressed by FGF2 and *fgfr2* gain-of-function mutations. **a**, Western analysis of BMP4-induced (concentrations are shown in ng ml^{-1}) *Noggin* in primary calvarial osteoblasts. **b**, FGF2 western analysis of cultured posterior frontal (PF), sagittal (SAG) and coronal (COR) dura mater. **c**, Western analysis of primary calvarial osteoblasts treated with BMP4 and FGF2 (concentrations of both are shown in ng ml^{-1}). **d**, *Noggin* western analysis of cultured posterior frontal (PF), sagittal (SAG) and coronal (COR) dura mater. Minus signs, untreated dura mater; plus signs, treatment with 100 ng ml^{-1} BMP4. **e**, *Noggin* western analysis of cultured coronal dura mater after treatment with FGF2 (concentrations shown in ng ml^{-1}) or transfection with an adenovirus containing *lacZ* (*AdlacZ*; 100 (++) multiplicity of infection (MOI)) or *fgfr2* (*AdCAsfgfr2*; 50 (+) and 100 (++) MOI, respectively). **f**, *Noggin* western analysis on sagittal dura mater transduced with a retrovirus containing wild-type *fgfr2* (WT *fgfr2*), Apert mutated *fgfr2* (S252W/*fgfr2*) or Crouzon mutated *fgfr2* (C342Y/*fgfr2*). **g**, *Noggin* western analysis on BMP4 (25 ng ml^{-1})-treated osteoblasts transduced with a retrovirus containing wild-type *fgfr2* (WT *fgfr2*), Apert mutated *fgfr2* (S252W/*fgfr2*) or Crouzon mutated *fgfr2* (C342Y/*fgfr2*).

coronal sutures) might not suppress BMP-induced Noggin expression, but environments high in FGF2 (that is, posterior frontal suture) reduce Noggin expression and enable suture fusion. These findings *in vitro* correlate with murine expression patterns *in vivo* for FGF2 (ref. 9) and Noggin in posterior frontal, sagittal and coronal sutures (Fig. 1c).

Whereas BMP4 stimulates Noggin expression in calvarial osteoblasts, the suture-specific dura mater is an independent source of Noggin; thus, the dura mater might be directly patterning cranial suture fate by interrupting BMP signalling. Cultured dura mater cells from patent sutures expressed high levels of Noggin protein without BMP stimulation, whereas the dura mater from the fusing posterior frontal suture expressed almost undetectable levels of Noggin. Treating these dural cell cultures with BMP4 had little impact on Noggin protein production (Fig. 2d), indicating that Noggin expression might be near maximal in sagittal and coronal dural cell cultures. Furthermore, treatment with BMP4 does not overcome endogenous FGF2-mediated Noggin suppression in the posterior frontal dura mater (Fig. 2d). Interestingly, when signalling mediated by FGF receptors (FGFRs) was disrupted with a dominant-negative FGFR1 adenoviral construct, posterior frontal dural cells did not express Noggin (data not shown). However, complete abrogation of FGF2 signalling is difficult, and the failure of our dominant-negative construct to restore Noggin expression might indicate that levels of the truncated FGFR1 receptor were insufficient to competitively inhibit wild-type receptors.

While blocking FGF signalling²⁰ or FGF2 activity²¹ prevents cranial suture fusion or osteogenesis, respectively, our findings suggest that exogenous FGF signalling is capable of suppressing Noggin expression during cranial suture fusion. For example, treatment with FGF2 suppressed Noggin production from coronal (Fig. 2e) and sagittal (data not shown) dura mater in a dose-dependent fashion. Taken together, these findings suggest that FGF2 guides suture fate (patency versus fusion) by regulating suture-specific Noggin production in osteoblasts and dura mater and, in turn, suture-specific BMP activity. Moreover, these data suggest a possible mechanism for syndromic craniosynostoses arising from *fgfr* gain-of-function mutations.

Because constitutive FGFR signalling is associated with syndromic forms of premature cranial suture fusion, we investigated the role of Noggin in a previously established model of FGF-mediated coronal synostosis²⁰. In this model, injection of an FGF2-expressing adenovirus into perinatal coronal dura mater led to FGF2 over-

expression, pathological osteogenesis and suture fusion within 30 days. Just as the application of FGF2 protein suppressed Noggin production in coronal dura mater in culture, infection with the FGF2-expressing adenovirus completely ablated Noggin expression (Fig. 2e). Injection of this FGF2-expressing adenovirus into the coronal dura mater of neonatal *lacZ/noggin* transgenic mice led to the suppression of Noggin expression in all animals (Fig. 3) and pathological coronal suture fusion (ref. 20, and data not shown). Taken together, our cell culture and *in vivo* data suggest that increased FGFR signalling might lead to suture fusion by suppressing Noggin production in the dura mater and osteoblasts of normally patent cranial sutures.

Because FGF2 misexpression led to Noggin suppression in coronal sutures, we investigated the effects of Apert (S252W) and Crouzon (C342Y) syndrome *fgfr2* gain-of-function mutations on Noggin production in dural cell and osteoblast cultures²². Although infection with a wild-type *fgfr2* retrovirus did not substantially affect Noggin expression, both Apert and Crouzon syndrome *fgfr2* mutants markedly downregulated Noggin protein production in sagittal dura mater (Fig. 2f). The Apert and Crouzon *fgfr2* constructs also downregulated BMP4-induced Noggin expression in calvarial osteoblasts (Fig. 2g). Because both Apert and Crouzon syndrome *fgfr* gain-of-function mutations promote pathological suture fusion, these findings provide an important link between the murine models and the gain-of-function *fgfr* mutations associated with syndromic forms of human craniosynostosis.

Having demonstrated that Noggin is normally expressed in the patent suture complex and that posterior frontal dura mater-derived FGF2 suppresses Noggin, we proposed that forced

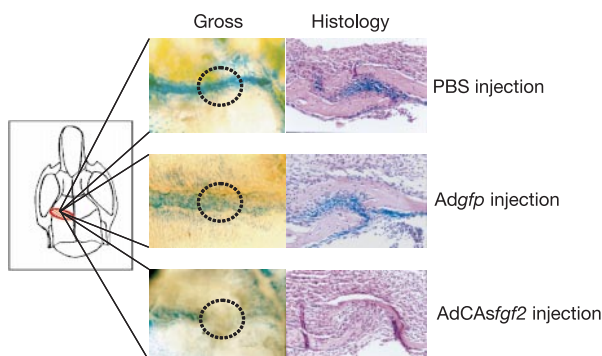


Figure 3 FGF2 suppresses *noggin* expression in coronal dura mater *in vivo*. Gross ($\times 5$ original magnification) and histological ($\times 100$ original magnification) views of stained *lacZ/noggin* transgenic 9-day-old coronal sutures after microinjection of PBS, 10^9 PFU of an adenovirus containing *gfp* (Adgfp) or 10^9 PFU of an adenovirus containing *fgf2* (AdCAsfgf2). Animals were microinjected on day 3 of life. Injection of PBS ($n = 9$) or Adgfp ($n = 11$) did not affect *noggin* expression. In marked contrast, AdCAsfgf2 ($n = 13$) led to *noggin* suppression in all animals transfected. The dotted circle denotes area of microinjection.

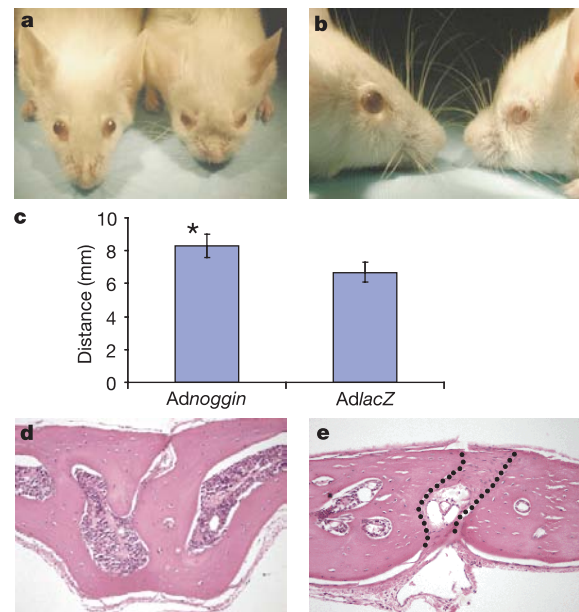


Figure 4 Noggin misexpression maintains posterior frontal suture patency *in vivo*. **a**, Anterior view of 53-day-old CD-1 mice. The posterior frontal sutures of these mice were injected with 10^9 PFU of an adenovirus containing *lacZ* (left, $n = 12$) or *noggin* (right, $n = 9$) on day 3 of life. **b**, Lateral view of mice injected with AdlacZ (left) and Adnoggin (right). Adnoggin-injected animals had a marked reduction in midface projection. **c**, The distance between the eyes of Adnoggin-injected mice was significantly greater than that of AdlacZ-injected mice (intercanthal distance 8.28 ± 0.7 mm versus 6.68 ± 0.6 mm, $*P < 0.0001$) owing to increased frontal bone growth perpendicular to the posterior frontal suture. **d**, Histological section ($\times 100$ original magnification) of the posterior frontal suture of AdlacZ-injected mouse. The posterior frontal suture is fused. **e**, Histological section ($\times 100$ original magnification) of the posterior frontal suture of an Adnoggin-injected mouse. Dotted lines indicate osteogenic fronts of widely patent posterior frontal suture.

expression of *noggin* would maintain posterior frontal suture patency. We first examined the effects of *noggin* misexpression in a previously established *in vitro* organ culture model²³. Using a *noggin*-expressing adenovirus, 22-day-old posterior frontal sutures were infected and placed in organ culture. After 30 days, all posterior frontal sutures mock-infected or infected with *lacZ* virus were fused. In marked contrast, all posterior frontal sutures infected with the *noggin* virus were widely patent (Supplementary Information).

To test the effects of *noggin* misexpression *in vivo*, the posterior frontal sutures of 3-day-old CD-1 mice were injected with PBS, a *lacZ* adenovirus (*AdlacZ*) or a *noggin* adenovirus (*Adnoggin*). After 50 days, the *Adnoggin*-infected mice had short broad snouts and widely spaced eyes (Fig. 4a and b). The distance between the eyes of *Adnoggin*-infected mice was significantly greater than that of *AdlacZ*-injected mice (intercanthal distance 8.28 ± 0.7 mm versus 6.68 ± 0.6 mm, $P < 0.0001$; Fig. 4c) owing to increased frontal bone growth perpendicular to the posterior frontal suture. Histological analysis demonstrated that all mice injected with PBS ($n = 11$) and *AdlacZ* virus ($n = 12$) had normally fused posterior frontal sutures, whereas the posterior frontal sutures of all *Adnoggin*-injected mice ($n = 9$) were widely patent (Fig. 4d and e). This result demonstrates that *noggin* misexpression at an early stage of suture development has profound consequences on cranial suture fate.

Homozygous *noggin*^{-/-} mice die at birth with multiple defects that include joint fusion of the appendicular skeleton⁶. While human *noggin* mutations have been linked to the fusion of finger joints (proximal symphalangism)²⁴, nothing was previously known about the role of *noggin* in calvarial joint (suture) biology until the opposing activity of Noggin and BMP4 in the dorsal lip of amphibian gastrulae led us to examine whether Noggin antagonized BMP4 in postnatal cranial sutures¹⁴. Here we have demonstrated that Noggin, a high-affinity secreted BMP antagonist, is present only in patent sutures and is decisively regulated by FGF signalling. Because Noggin enforces suture patency, syndromic *fgfr*-mediated craniosynostoses might result from inappropriate Noggin suppression. Finally, ectopic *noggin* expression prevents the fusion of mouse posterior frontal sutures, raising the possibility that therapeutic Noggin could be exploited to control postnatal skeletal development. □

Methods

Gene chip analysis

Total RNA was isolated and pooled from posterior frontal and sagittal sutures ($n = 20$ per time point) with associated dura mater from Sprague–Dawley rats at times before, during and after programmed posterior frontal suture fusion. Synthesis of cDNA, *in vitro* transcription and biotin labelling of cRNA, and hybridization to the R34U A Chips (Affymetrix) were performed in accordance with Affymetrix protocols. Expression data were analysed with Microarray Suite 5.0 (Affymetrix) and Genespring 4.2 (Silicon Genetics) software. Overall chip intensities were scaled to an artificial mean to permit chip-to-chip comparisons and normalized to day 30 (post-fusion) sutures.

Quantitative real-time RT–PCR

RNA was harvested with the Ambion RNAqueous for PCR kit in accordance with the manufacturer's instructions, including incubation with DNase I to remove any genomic contamination. Taqman (Applied Biosystems, Foster City, California) primer and probe sets for *bmp2*, *bmp4* and *bmp7* were designed with PrimerExpress Software (Applied Biosystems) and screened for appropriate amplification by using end-point RT–PCR analysis. Rodent glyceraldehyde-3-phosphate dehydrogenase (*gapdh*) primer and probe reagents were purchased from Applied Biosystems. Quantification was performed with the ABI Prism 7900HT Sequence Detection System (Applied Biosystems) in a two-step non-multiplexed assay in accordance with the relative standard method. Transcript quantification was performed in triplicate for every sample and reported relative to *gapdh*.

Immunohistochemistry

Immunohistochemistry was performed on 18-day-old mice as described previously²⁵. Sections were de-waxed in xylene, blocked in 10% goat serum, incubated in polyclonal BMP4 antibody (R&D Systems) or PBS alone (control), followed by a biotinylated anti-rabbit antibody, and amplified with an avidin–biotin enzyme complex (ABC kit; Vector Laboratories). Detection was performed with 0.05% 3,3'-diaminobenzidine chromophore (Sigma) and 0.03% hydrogen peroxide in 0.1 M Tris-HCl, pH 7.4. Sections were lightly counterstained with haematoxylin.

In situ hybridization

Calvaria were harvested, decalcified and embedded as described previously²⁶. Antisense *gremlin*, *dcf7*, *prdc*, *sost*, *dcf7*, *cerberus* and *noggin* riboprobes were labelled with ³⁵S-UTP and *in situ* hybridization was performed with standard protocols²⁷. Sections were counterstained with Hoechst 33258 ($5 \mu\text{g ml}^{-1}$) to reveal nuclei.

LacZ staining

Calvaria were stained as described previously⁶. Whole calvaria were fixed in 4% paraformaldehyde for 20 min and placed in standard staining solution (5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, 2 mM MgCl₂, 0.4% 5-bromo-4-chloro-4-indolyl-β-D-galactopyranoside in PBS) overnight at 4 °C. The following day, samples were incubated for 4–6 h at 37 °C (wild-type littermates served as controls). Stained calvaria were dehydrated through an alcohol series and embedded in paraffin. Sections were cut at 5 μm thickness and lightly counterstained with haematoxylin and eosin.

Western blot analysis

Dura mater or osteoblast cultures were grown to confluence, starved in low-serum medium for 24 h and then treated with various doses of BMP4 and/or FGF2. Medium was collected 24 h after cytokine treatment and incubated with immobilized heparin–Sepharose beads (Pharmacia). Bound proteins were eluted by heating at 95 °C. Denatured samples were fractionated by 10% SDS–polyacrylamide-gel electrophoresis, transferred to a poly(vinylidene difluoride) membrane and probed with RP57–16 anti-Noggin antibody²⁸. Blots were exposed to a horseradish-peroxidase-conjugated goat anti-rat IgG and developed with ECL chemiluminescent reagents (Amersham Pharmacia Biotech). All membranes were stained with Ponceau S to ensure equivalent loading and transfer.

Retroviral infection

The pLSXN/*fgfr2* constructs were co-transfected with an ecotropic packaging vector into HEK-293 cells. Supernatants were collected, passed through a 0.45 μm filter and frozen at –80 °C. Osteoblasts and dural cells were infected in the presence of $8 \mu\text{g ml}^{-1}$ Polybrene (Sigma) and selected in medium containing G418 ($400 \mu\text{g ml}^{-1}$).

Adenoviral infection in vitro

Adenoviruses were prepared as previously described²⁰. Calvaria from 22-day-old CD-1 mice were isolated and washed. The posterior frontal sutures were injected with PBS ($n = 6$), *AdlacZ* (10^9 plaque-forming units (PFU), $n = 7$) or *Adnoggin* (10^9 PFU, $n = 5$). Infected and control calvaria were placed on a 3.0-μm culture well insert and fed with standard BGJb medium (pH 7.4). All sutures were harvested after 30 days in culture and immediately fixed in 4% paraformaldehyde. Calvaria were photographed, processed, embedded and sectioned.

Adenoviral infection in vivo

To determine the effects of FGF2 on *noggin* expression, the coronal sutures of 3-day-old transgenic *lacZ/noggin* mice were injected with PBS ($n = 9$), an adenovirus containing green fluorescent protein (*Adgfp*, 10^9 PFU, $n = 11$) or an adenovirus containing *fgf2* (*AdCasfgf2*, 10^9 PFU, $n = 13$). Animals were harvested 6 days after injection.

To determine the effects of *noggin* expression on suture patency, the posterior frontal sutures of 3-day-old CD-1 mice were injected with PBS ($n = 11$), an adenovirus containing *lacZ* (10^9 PFU, $n = 12$) or an adenovirus containing *noggin* (10^9 PFU, $n = 9$). Previous work has documented widespread infection by such viruses, with expression detectable for 30 days (ref. 20). At 50 days after injection, animals were harvested and photographed, and intercanthal distances were measured with digital calipers with an accuracy of 0.01 mm (Model CD-6 CS, Mitutoyo, Kanagawa, Japan). A paired *t*-test was performed to determine statistical significance. After measurement, the skulls were processed, embedded and sectioned.

Received 2 October 2002; accepted 7 March 2003; doi:10.1038/nature01545.

- Slavkin, H. C. *Developmental Craniofacial Biology* (Lea & Febiger, Philadelphia, 1979).
- Thilander, B. Basic mechanisms in craniofacial growth. *Acta Odontol. Scand.* **53**, 144–151 (1995).
- McCarthy, J. G., Epstein, F. J. & Wood-Smith, D. in *Plastic Surgery* (ed. McCarthy, J. G.) 3013–3053 (W.B. Saunders Co., Philadelphia, 1990).
- Cohen, M. M. Jr Craniosynostoses: phenotypic/molecular correlations. *Am. J. Med. Genet.* **56**, 334–339 (1995).
- Wilkie, A. O. M. Craniosynostosis: genes and mechanisms. *Hum. Mol. Genet.* **6**, 1647–1656 (1997).
- Brunet, L. J., McMahon, J. A., McMahon, A. P. & Harland, R. M. Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. *Science* **280**, 1455–1457 (1998).
- McMahon, J. A. et al. Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. *Genes Dev.* **12**, 1438–1452 (1998).
- Capdevila, J. & Johnson, R. L. Endogenous and ectopic expression of *noggin* suggests a conserved mechanism for regulation of BMP function during limb and somite patterning. *Dev. Biol.* **197**, 205–217 (1998).
- Warren, S. M. et al. New developments in cranial suture research. *Plast. Reconstr. Surg.* **107**, 523–540 (2001).
- Opperman, L. A., Nolen, A. A. & Ogle, R. C. TGF-β1, TGF-β2, and TGF-β3 exhibit distinct patterns of expression during cranial suture formation and obliteration *in vivo* and *in vitro*. *J. Bone Miner. Res.* **12**, 301–310 (1997).
- Greenwald, J. A. et al. Regional differentiation of cranial suture-associated dura mater *in vivo* and *in vitro*: implications for suture fusion and patency. *J. Bone Miner. Res.* **15**, 2413–2430 (2000).
- Wozney, J. M. et al. Novel regulators of bone formation: Molecular clones and activities. *Science* **242**, 1528–1534 (1988).
- Wang, E. A. et al. Recombinant human bone morphogenetic protein induces bone formation. *Proc. Natl Acad. Sci. USA* **87**, 2220–2224 (1990).

14. Zimmerman, L. B., De Jesus-Escobar, J. M. & Harland, R. M. The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell* **86**, 599–606 (1996).
15. Hsu, D. R., Economides, A. N., Wang, X., Eimon, P. M. & Harland, R. M. The *Xenopus* dorsalizing factor gremlin identifies a novel family of secreted proteins that antagonize BMP activities. *Mol. Cell* **1**, 673–683 (1998).
16. Capdevila, J., Tsukui, T., Rodriguez Esteban, C., Zappavigna, V. & Izpisua Belmonte, J. C. Control of vertebrate limb outgrowth by the proximal factor Meis2 and distal antagonism of BMPs by gremlin. *Mol. Cell* **4**, 839–849 (1999).
17. Minabe-Saegusa, C., Saegusa, H., Tsukahara, M. & Noguchi, S. Sequence and expression of a novel mouse gene PRDC (protein related to DAN and cerberus) identified by a gene trap approach. *Dev. Growth Differ.* **40**, 343–353 (1998).
18. Brunkow, M. E. *et al.* Bone dysplasia sclerosteosis results from loss of the SOST gene product, a novel cysteine knot-containing protein. *Am. J. Hum. Genet.* **68**, 577–589 (2001).
19. Gazzerro, E., Gangji, V. & Canalis, E. Bone morphogenetic proteins induce the expression of noggin, which limits their activity in cultured rat osteoblasts. *J. Clin. Invest.* **102**, 2106–2114 (1998).
20. Greenwald, J. A. *et al.* *In vivo* modulation of FGF biological activity alters cranial suture fate. *Am. J. Pathol.* **158**, 441–452 (2001).
21. Moore, R., Ferretti, P., Copp, A. & Thorogood, P. Blocking endogenous FGF-2 activity prevents cranial osteogenesis. *Dev. Biol.* **243**, 99–114 (2002).
22. Mansukhani, A., Bellota, P., Sahni, M. & Basilico, C. Signaling by fibroblast growth factors (FGF) and fibroblast growth factor receptor 2 (FGFR2)-activating mutations blocks mineralization and induces apoptosis in osteoblasts. *J. Cell Biol.* **149**, 1297–1308 (2000).
23. Bradley, J. P. *et al.* Studies in cranial suture biology: *in vitro* cranial suture fusion. *Cleft Palate–Craniofac. J.* **33**, 150–156 (1996).
24. Dixon, M. E., Armstrong, P., Stevens, D. B. & Bamshad, M. Identical mutations in NOG can cause either tarsal/carpal coalition syndrome or proximal symphalangism. *Gen. Med.* **3**, 349–353 (2001).
25. Roth, D. A. *et al.* Studies in cranial suture biology. I. Increased immunoreactivity for transforming growth factor- β (β 1, β 2, β 3) during rat cranial suture fusion. *J. Bone Miner. Res.* **12**, 311–321 (1997).
26. Bradley, J. P., Levine, J. P., Roth, D. A., McCarthy, J. G. & Longaker, M. T. Studies in cranial suture biology. IV. Temporal sequence of posterior frontal cranial suture fusion in the mouse. *Plast. Reconstr. Surg.* **98**, 1039–1045 (1996).
27. Albrecht, U., Helms, J. A. & Lin, H. In *Molecular and Cellular Methods in Developmental Toxicology* (ed. Daston, G. P.) 23–48 (CRC Press, Boca Raton, 1997).
28. Paine-Saunders, S., Viviano, B. L., Economides, A. N. & Saunders, S. Heparan sulfate proteoglycans retain noggin at the cell surface: A potential mechanism for shaping bone morphogenetic protein gradients. *J. Biol. Chem.* **277**, 2089–2096 (2002).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank C. J. Tabin, I. Thesleff, D. M. Kingsley and N. Quarto for comments and suggestions on this manuscript, and K. D. Fong, J. A. Mathy and R. P. Nacamuli for their technical assistance. The human *fgfr2*, *fgfr2/S252W* (Apert) and *fgfr2/C342Y* (Crouzon) retroviral constructs were provided by A. Mansukhani and C. Basilico (New York University). This work was supported by NIH grants (R.M.H. and M.T.L.) and a Lyndon Peer/PSEF Fellowship (S.M.W.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.T.L.
(e-mail: longaker@stanford.edu).

The exocyst complex is required for targeting of Glut4 to the plasma membrane by insulin

Mayumi Inoue, Louise Chang, Joseph Hwang, Shian-Huey Chiang & Alan R. Saltiel

Life Sciences Institute, Departments of Internal Medicine and Physiology, University of Michigan Medical Center, Ann Arbor, Michigan 48109, USA

Insulin stimulates glucose transport by promoting exocytosis of the glucose transporter Glut4 (refs 1, 2). The dynamic processes involved in the trafficking of Glut4-containing vesicles, and in their targeting, docking and fusion at the plasma membrane, as well as the signalling processes that govern these events, are not well understood. We recently described tyrosine-phosphorylation events restricted to subdomains of the plasma membrane that result in activation of the G protein TC10 (refs 3, 4). Here we

show that TC10 interacts with one of the components of the exocyst complex, Exo70. Exo70 translocates to the plasma membrane in response to insulin through the activation of TC10, where it assembles a multiprotein complex that includes Sec6 and Sec8. Overexpression of an Exo70 mutant blocked insulin-stimulated glucose uptake, but not the trafficking of Glut4 to the plasma membrane. However, this mutant did block the extracellular exposure of the Glut4 protein. So, the exocyst might have a crucial role in the targeting of the Glut4 vesicle to the plasma membrane, perhaps directing the vesicle to the precise site of fusion.

To search for potential effectors of TC10 that have a role in insulin-stimulated glucose transport, we screened a yeast two-hybrid complementary-DNA library derived from 3T3L1 adipocytes with a constitutively active mutant form of human TC10 α ⁵, and identified two clones encoding the full-length sequence of Exo70, a component of the exocyst complex, which has been implicated in the tethering or docking of secretory vesicles^{6,7}. To explore the specificity of the interaction of TC10 with Exo70, we co-transfected 293T cells with myc-tagged Exo70 and haemagglutinin (HA)-tagged, constitutively active forms of mouse TC10 α , Rho1, Rac1 or Cdc42 cDNAs^{8–11} (Fig. 1a). Exo70 could be co-precipitated with TC10, but not with Rho, Rac or Cdc42, indicating that its interaction with TC10 is specific. TC10 β could also be co-precipitated with Exo70 (data not shown). To evaluate whether Exo70 is a potential effector of TC10, we incubated lysates derived from cells expressing different forms of HA-tagged TC10 with a glutathione S-transferase (GST)–Exo70 fusion protein, to examine binding *in vitro* (Fig. 1b). Exo70 bound to the constitutively active mutant TC10 α (Q67L), but not to the dominant-interfering TC10 α (T23N). Identical results were obtained with the corresponding mutants TC10 β (T25N) and TC10 β (Q69L) (data not shown). To confirm that the interaction of TC10 with Exo70 was GTP-dependent, lysates expressing TC10 were loaded with GDP or a non-hydrolysable form of GTP (GTP γ S) *in vitro*, and incubated with GST–Exo70. Exo70 bound weakly to the GDP-bound form of TC10, and strongly to the GTP-bound, activated state of the protein. These data show that Exo70 interacts with TC10 in a GTP-dependent manner.

To study the function of Exo70 in glucose transport, we sought to map its TC10 binding site through a series of deletion mutants. Exo70 has only one putative protein–protein interaction region, a coiled-coil domain in the amino terminus (Fig. 1c). We deleted amino acids 1–384 of Exo70, and also prepared constructs expressing only amino acids 1–384, 1–99 or 100–384, and evaluated the ability of the mutants to bind constitutively active TC10 α (Q67L) in a co-precipitation experiment. Whereas wild-type Exo70 (Exo70-WT) bound to TC10, Exo70(385–653) (Exo70-C) did not. Moreover, an Exo70 fragment containing only amino acids 1–384 (Exo70-N) bound to TC10, whereas Exo70 containing amino acids 1–99, which comprised the coiled-coil domain, bound weakly, and Exo70 containing residues 100–384 also bound weakly. These data indicate that amino acids 1–384 of Exo70 are both necessary and sufficient for binding to TC10, but that this interaction requires more than the coiled-coil domain (Fig. 1d and Supplementary Fig. 1).

We analysed the intracellular localization of Exo70 by immunohistochemistry, and evaluated its potential regulation by insulin and TC10. 3T3L1 adipocytes were transfected with myc–Exo70 plus or minus TC10 and its mutants. As shown in Fig. 2a, b, myc–Exo70 was distributed diffusely throughout the intracellular region. Co-expression of TC10 or its inactive mutant TC10(T23N) did not influence the localization of overexpressed myc–Exo70. By contrast, Exo70 was localized at the plasma membrane when co-expressed with the constitutively active mutant TC10(Q67L), indicating that TC10 activation stimulates the translocation of Exo70. We also examined whether the amino-terminal fragment of Exo70 (Exo70-N) translocated to the plasma membrane after expression of active TC10. Unlike the wild-type form of the protein, this fragment did

not translocate to the membrane when co-expressed with active TC10, despite the fact that it can bind to the protein *in vitro* or after co-expression in 293T cells (Fig. 1d). These data suggest that the carboxy-terminal sequences in Exo70 are required for the movement of the protein in adipocytes, possibly owing to an interaction with another protein.

The translocation of Exo70 to the plasma membrane by activated TC10 suggested that insulin might induce a similar redistribution of the protein. Following transfection with myc-Exo70, cells were stimulated with insulin, and the localization of the protein was examined by immunohistochemistry. Insulin stimulation resulted in the translocation of Exo70 to the plasma membrane. This event was blocked by the co-transfection of cells with the inactive mutant TC10(T23N) (Fig. 2c, d). By contrast, co-expression with the wild-type or constitutively active mutant of TC10 did not block the effect of insulin (data not shown). Interestingly, as observed above with the overexpression of active TC10, insulin failed to stimulate the translocation of the amino-terminal fragment of Exo70 to the plasma membrane.

The observation that insulin can recruit Exo70 to the plasma membrane by way of TC10 suggests that insulin might promote the assembly of the exocyst complex through an interaction between TC10 and Exo70. The exocyst complex is thought to assemble before the fusion of secretory vesicles with the plasma membrane^{12–14}. This complex includes the proteins Sec6 and Sec8, which have been shown to regulate the late stage of exocytosis in MDCK cells^{15,16}. Unfortunately, antibodies to Sec6 and Sec8 could not be used for

immunohistochemistry. So, to evaluate the localization of these endogenous proteins, we fractionated cell lysates derived from 3T3L1 adipocytes stimulated with or without insulin. Sec6 and Sec8 were found in cytosolic and low-density microsomal (LDM) fractions, with a small amount found in the plasma-membrane fraction. Treatment of cells with insulin reduced the appearance of the two proteins in the cytoplasm and low-density microsomes, with a corresponding increase in the plasma-membrane fraction (Fig. 3a).

We also evaluated whether Exo70 formed a complex with Sec6 and Sec8. As shown in Fig. 3b, endogenous Sec6 bound to Sec8, as assayed by co-precipitation. This interaction was insulin-indepen-

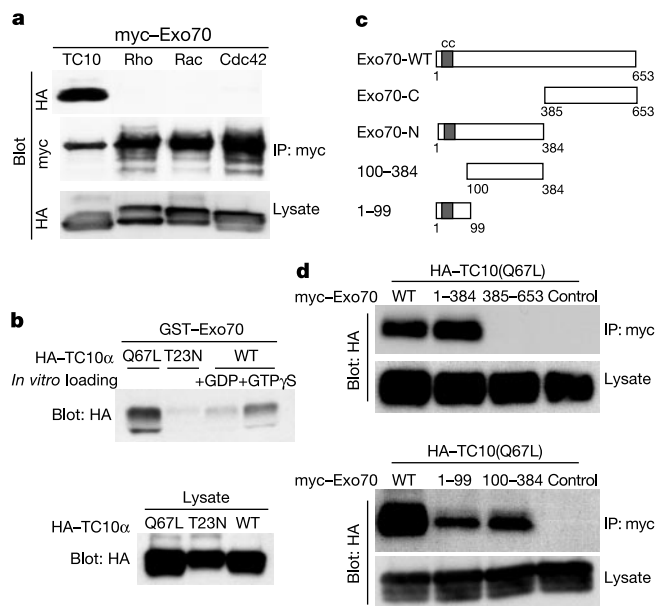


Figure 1 Exo70 interacts with activated TC10. **a**, Exo70 specifically interacts with TC10. 293T cells were co-transfected with myc-Exo70 and HA-tagged, constitutively active forms of TC10α(Q67L), RhoA(G14V), Rac1(G12V) and Cdc42(Q61L). Lysates were immunoprecipitated (IP) with an anti-myc antibody, followed by blotting with anti-HA antibody. **b**, Exo70 interacts with TC10 in a GTP-dependent manner. 293T cells were transfected with HA-tagged wild-type TC10α, its constitutively active Q67L mutant, or its inactive T23N mutant. Cell lysates were preincubated with either 100 μM GTPγS or 1 mM GDP, and then incubated with glutathione-Sepharose-bound GST-Exo70. Precipitates were analysed by immunoblot analysis with anti-HA antibody. Identical results were obtained with TC10β (data not shown). **c**, Schematic of Exo70 and its deletion mutants. cc, coiled-coil. **d**, Interaction of TC10 with Exo70 mutants. 293T cells were co-transfected with HA-TC10α(Q67L) and myc-tagged wild-type Exo70 or mutants containing amino acids 1–384 (Exo70-N), 385–653 (Exo70-C), 1–99, 100–384 or empty vector. Lysates were immunoprecipitated with anti-myc antibody, and immunoprecipitates were analysed by blotting with anti-HA antibody.

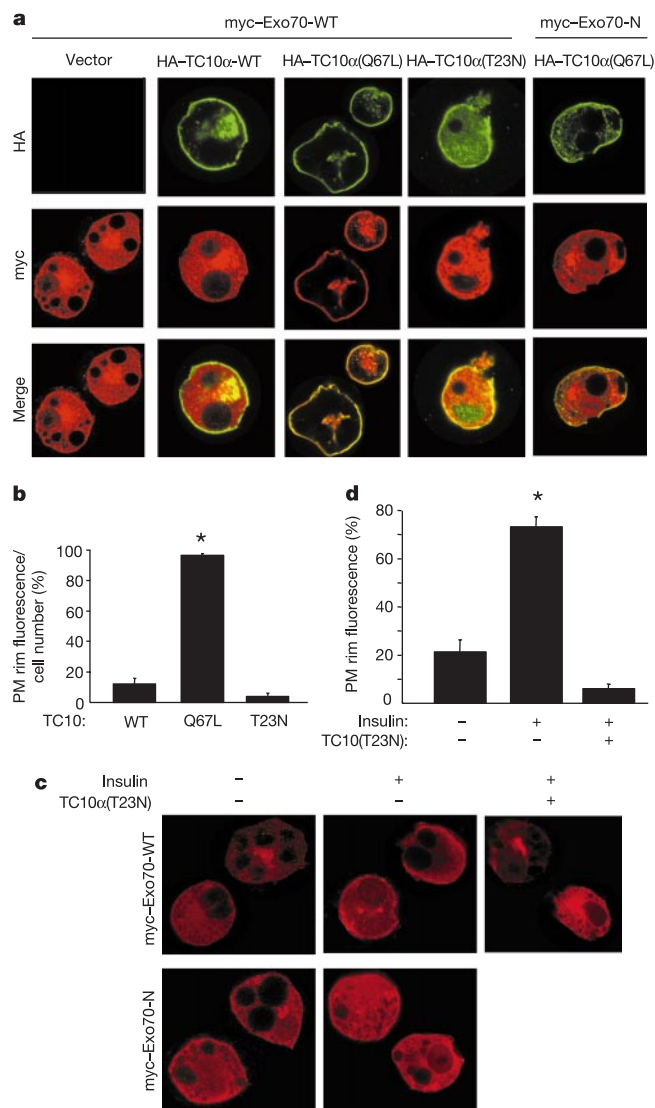


Figure 2 Exo70 is translocated to the plasma membrane by insulin. **a**, Overexpression of constitutively active TC10(Q75L) recruits Exo70 to the plasma membrane (PM). 3T3L1 adipocytes were co-transfected with myc-Exo70-WT or myc-Exo70-N with vector only, HA-tagged wild-type TC10, or its Q75L or T31N mutants. Cells were fixed and stained with anti-myc polyclonal antibody to detect myc-Exo70, and with anti-HA monoclonal antibody to detect HA-TC10, followed by Alexa⁵⁹⁴ goat anti-rabbit IgG and Alexa⁴⁸⁸ goat anti-mouse IgG. Fluorescence was visualized by confocal laser scanning microscopy. Images represent four independent determinations. **b**, **d**, Graphical representation of data from three independent experiments. *Significant difference, $P < 0.0001$. **c**, Exo70 translocates to the plasma membrane in response to insulin. 3T3L1 adipocytes were electroporated with myc-Exo70-WT or myc-Exo70-N with or without HA-TC10(T31N), serum-starved, and treated with or without insulin for 5 min. Cells were fixed and stained with anti-myc antibody followed by Alexa⁵⁹⁴ goat anti-mouse IgG.

dent. Moreover, both wild-type Exo70 and its amino-terminal fragment bound to endogenous Sec8 (Fig. 3c, d). Interestingly, whereas the amino-terminal region of Exo70 bound to Sec8, a fragment of the protein containing only amino acids 385–653 did not (data not shown). So, Exo70 appears to exist in a preformed complex with Sec6 and Sec8, through an interaction in the amino-terminal region of Exo70. This multiprotein complex is then translocated to the plasma membrane in response to insulin.

As the amino-terminal fragment of Exo70 fails to translocate to the plasma membrane in response to insulin or active TC10, this mutant form of the protein might also block the assembly of the exocyst complex at the cell surface. Cells were transfected with the myc-tagged wild-type or amino-terminal fragment of Exo70 (Fig. 3e), treated with insulin, and fractionated as detailed above. As determined by immunohistochemical analysis (Fig. 2), insulin stimulated the translocation of wild-type Exo70 to the plasma membrane. However, the amino-terminal fragment remained mainly in the cytoplasm, and the localization of this protein was not influenced by insulin. Sec8 also translocated to the membrane in response to insulin in cells that expressed wild-type Exo70.

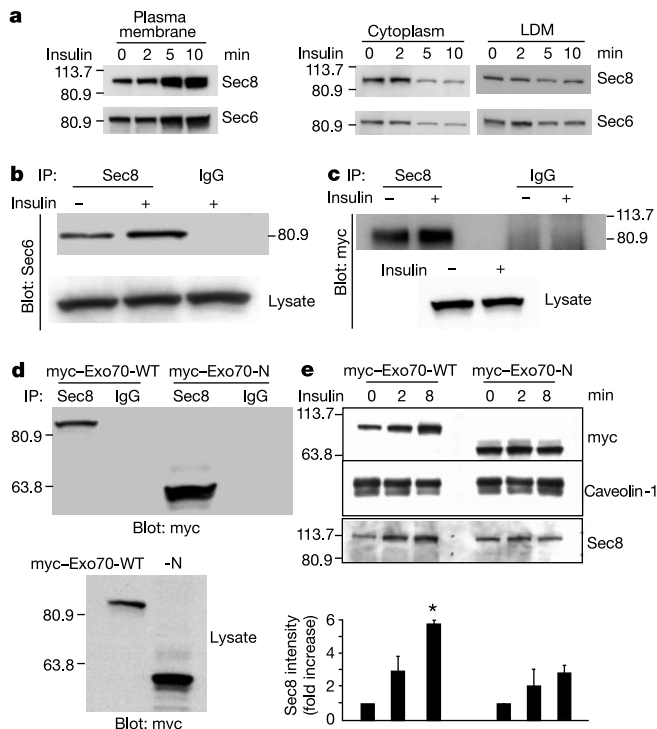


Figure 3 Components of the exocyst complex translocate to the plasma membrane in response to insulin. **a**, Sec6 or Sec8 translocate to the plasma membrane in response to insulin stimulation. After insulin stimulation, cells were placed on ice and washed three times with ice-cold PBS. Membrane fractionation was carried out using 3T3L1 adipocytes as reported²⁹, and endogenous Sec6 and Sec8 were detected by monoclonal antibodies. **b**, Sec8 interacts with Sec6. The cell lysates from 3T3L1 adipocytes were incubated with a Sec8 monoclonal antibody and blotted with a Sec6 antibody. **c**, Sec8 and Exo70 interact. 3T3L1 adipocytes were transfected with myc-Exo70. Cells were stimulated with 100 nM insulin for 10 min. Cell lysates were incubated with a Sec8 monoclonal antibody and blotted with an anti-myc polyclonal antibody. **d**, Sec8 interacts with Exo70-WT and Exo70-N. Lysates were prepared as described in **c** from 3T3L1 adipocytes transfected with myc-Exo70-WT or myc-Exo70-N. **e**, Exo70-N blocks membrane translocation of Sec8. Membrane fractionation was carried out using 3T3L1 adipocytes transfected with myc-Exo70-WT and myc-Exo70-N after insulin stimulation. Endogenous Sec8 and caveolin-1 were detected by western blotting using the plasma-membrane fraction. The means (and standard deviations) of three independent determinations are shown in graphical form. *Significant difference, $P < 0.005$.

However, expression of the mutant form of the protein blocked the translocation of endogenous Sec8. These data indicate that the amino-terminal fragment of Exo70 can act as a dominant-negative mutant, presumably blocking the insulin-stimulated assembly of the exocyst complex at the plasma membrane.

The exocyst complex is thought to have an important role in targeted exocytosis. To evaluate whether the interaction between TC10 and Exo70 is involved in Glut4 vesicle trafficking, we over-expressed Exo70 and its mutants in 3T3L1 adipocytes by electroporation, and assayed 2-deoxyglucose uptake¹⁷ (Fig. 4a). In cells transfected with an empty vector, insulin produced a fivefold increase in 2-deoxyglucose uptake. Overexpression of wild-type Exo70 led to a 48% increase in insulin-stimulated glucose uptake over that observed in cells transfected with an empty vector, although basal activity was not significantly changed. By contrast, the amino-terminal fragment of Exo70, which blocks the assembly of the exocyst complex, produced a 40% inhibition of insulin-stimulated transport compared with control, although basal activity was again unaffected. In these experiments, less than 50% of the cells were transfected by the electroporation method, indicating that

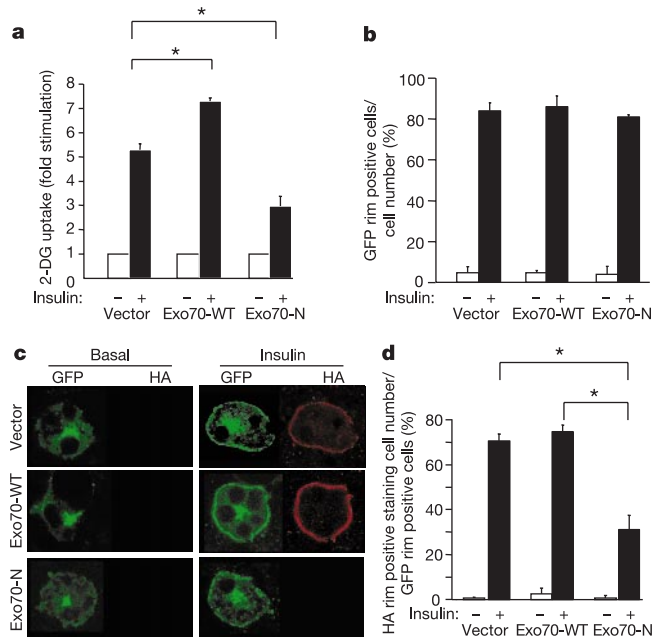


Figure 4 Exo70 plays a crucial role in Glut4 vesicle targeting. **a**, Differentiated 3T3L1 adipocytes were transfected with Exo70-WT or Exo70-N. The cells were left untreated or were stimulated with 100 nM insulin for 30 min. The rate of [¹⁴C]2-deoxyglucose (2-DG) uptake was determined. Cells were transfected with an efficiency of approximately 50%. Results are expressed as the mean $\pm$ s.d. of five separate experiments. *Significant difference, $P < 0.01$. **b**, Number of Glut4-eGFP transfected cells with visually detectable plasma-membrane rim fluorescence. The numbers of Glut4-eGFP-transfected cells displaying visually detectable eGFP rim fluorescence are plotted. **c**, Exo70-N does not affect HA-Glut4-eGFP translocation, but blocks its extracellular exposure. 3T3L1 adipocytes were co-electroporated with HA-Glut4-eGFP plus vector alone, myc-tagged Exo70-WT or Exo70-N. Cells were serum-starved, treated with or without insulin for 30 min, fixed, and stained with an anti-HA monoclonal antibody followed by Alexa⁵⁹⁴ goat anti-mouse IgG to detect cells that expressed HA-Glut4 at the surface. Fluorescence was visualized by confocal laser scanning microscopy. Images represent three independent determinations. **d**, Percentage of the HA rim positive staining cells among the HA-Glut4-GFP-transfected cells was quantified. Fifty cells that showed translocation of eGFP were scored for visually detectable HA rim fluorescence before and after treatment with insulin for 30 min. These data were obtained by blind counting of more than 50 cells from three independent experiments. *Significant difference, $P < 0.0001$.

most of the insulin-stimulated glucose transport was blocked by overexpression of the Exo70-N mutant.

As mentioned above, insulin-stimulated glucose transport results from the translocation of the Glut4 protein to the cell surface. So, we next evaluated the effect of Exo70 on the translocation of a Glut4-enhanced green fluorescent protein (eGFP) fusion protein in 3T3L1 adipocytes. Cells were co-transfected with Glut4-eGFP and with wild-type or mutant forms of Exo70. Insulin stimulated the translocation of Glut4-eGFP in 90% of the cells transfected with vector alone, as detected by increased rim fluorescence. Co-expression with wild-type Exo70 had no impact on basal or insulin-stimulated translocation. Interestingly, co-expression with the amino-terminal fragment of Exo70 was also without effect on Glut4-eGFP translocation, despite the clear inhibition of glucose transport observed by overexpression of this mutant (Fig. 4b).

The inhibition of insulin-stimulated glucose uptake in the face of no effect on Glut4 translocation by the Exo70 mutant indicated that the protein might influence the targeting or docking of Glut4 to the plasma membrane, rather than the translocation process. To explore this possibility in more detail, we co-expressed the Exo70 mutants with a Glut4-GFP construct that was HA-tagged in its first extracellular loop¹⁸ (Fig. 4c). This double-tagged form of Glut4 can be detected by the fluorescence of GFP, and simultaneously by immunodetection of the HA tag at the cell surface in non-permeabilized cells, thus representing a method to evaluate both the translocation and extracellular exposure of Glut4 (ref. 18). In the absence of insulin, there was no cell-surface staining of HA-Glut4-GFP with anti-HA, whereas insulin treatment produced a strong signal. Co-transfection of cells with wild-type Exo70 significantly increased the cell-surface staining of HA-Glut4-GFP, as detected by anti-HA antibodies, consistent with the increase in glucose transport produced by overexpression of the protein. By contrast, co-transfection of cells with the amino-terminal fragment of Exo70 completely blocked the appearance of HA-Glut4-GFP at the cell surface (Fig. 4c, d), indicating that Exo70 might have a crucial role in the membrane targeting of Glut4. An inhibitory mutant of TC10 also blocked the cell-surface exposure of exofacial tagged Glut4 (Supplementary Fig. 2).

The translocation of the insulin-responsive glucose transporter Glut4 from intracellular membranes to the cell surface is likely to involve numerous steps, typical of vesicular transport in other systems. These may include mechanisms for the tethering of the vesicle to target membranes, where it can come into contact with the docking and fusion machinery. For many exocytotic processes, this step is mediated by the exocyst, an evolutionarily conserved octameric complex that targets vesicles to specific sites of the plasma membrane¹⁹. Data presented here suggest that the exocyst complex assembles at the plasma membrane of fat cells in response to insulin, through the association of Exo70 with activated TC10, and has a role in the tethering of the Glut4 vesicle, where it can come into proximity with the SNARE (SNAP receptor) complexes involved in docking and fusion^{20,21}. The Rab proteins, which are mammalian homologues of the yeast exocyst component Sec4, might mediate this tethering step. Rab4 has been found in Glut4 vesicles, and translocates to the plasma membrane in response to insulin^{22–25}. Moreover, we have also shown that TC10 can interact with the effector protein CIP4, which is required for Glut4 translocation²⁶. So, TC10 might engage multiple effectors that influence both Glut4 translocation and tethering of the vesicle at the plasma membrane. This possibility is under investigation. □

Methods

Yeast two-hybrid screen

The cDNA encoding full-length human TC10(Q75L) was amplified by polymerase chain reaction (PCR) using the plasmid pKH3-TC10(Q75L) (a gift from I. Macara, University of Virginia, Charlottesville) as a template, and a yeast two-hybrid cDNA library derived from 3T3L1 adipocyte messenger RNA was screened as described previously²⁶.

Expression constructs

To generate a mammalian expression vector, full-length Exo70 cDNA was subcloned into the pCS2 vector. To construct deletion mutants, Exo70 cDNA was digested with *NcoI* and religated, generating Exo70-N or Exo70-C mutants. For 1–99 or 100–384 mutants, Exo70 was PCR-amplified from pCS2-Exo70 using oligonucleotides containing *EcoRI/XhoI* restriction sites, and subcloned into pCS2 at *EcoRI/XhoI*.

In vitro GST pull-down assay

GST or GST-Exo70 fusion proteins were expressed in the BL21(DE3) pLysS *Escherichia coli* strain and purified. Cell lysates were prepared for immunoprecipitation as described below, and incubated with either GST alone or GST-Exo70 bound to glutathione-Sepharose beads (Amersham Pharmacia) for 2 h at 4 °C. For GTP-binding experiments, cell lysates were loaded *in vitro* with either 1 mM GDP or 100 μM GTPγS for 30 min at 37 °C before incubating with GST fusion proteins²⁷. The beads were washed five times with lysis buffer, resuspended in sample buffer, subjected to SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and analysed by immunoblotting with anti-myc monoclonal antibody or polyclonal antibody (Santa Cruz Biotechnology) or anti-HA monoclonal antibody (Santa Cruz Biotechnology).

Immunoprecipitation and immunoblotting

Cell lysates were prepared as described using 1% Triton X-100 (ref. 4). Lysates were then incubated with the indicated antibodies for 2 h at 4 °C. The immune complexes were precipitated with protein A or G agarose (Amersham Pharmacia) for 1 h at 4 °C, washed extensively with lysis buffer, resolved in 10–20% gradient SDS-PAGE, and analysed by immunoblotting. All immunoblots were developed by enhanced chemiluminescence (Amersham Pharmacia). Anti-Sec6 (9H5) and Sec8 (2E9) monoclonal antibodies were obtained from Calbiochem, and anti-caveolin-1 polyclonal antibody was obtained from Transduction Laboratories.

Immunofluorescence microscopy

For immunofluorescence studies, electroporated cells were processed as described¹⁷. To detect myc-Exo70, cells were stained with anti-myc monoclonal antibody or polyclonal antibody (Santa Cruz Biotechnology). To detect HA-TC10, cells were stained with anti-HA monoclonal antibody (Santa Cruz Biotechnology). After incubation with primary antibodies, cells were incubated with Alexa⁴⁸⁸ or Alexa⁵⁹⁴ goat anti-mouse or anti-rabbit immunoglobulin (IgG), obtained from Molecular Probes. Images were captured by using an Olympus FV300 confocal laser scanning microscope.

Assays of Glut4-eGFP translocation and HA-Glut4-GFP

Glut4-eGFP translocation assays were performed as described²⁸. The HA-Glut4-GFP construct was a gift from S. W. Cushman (National Institutes of Health, Maryland). The method was followed as described previously¹⁸.

Received 25 November 2002; accepted 4 March 2003; doi:10.1038/nature01533.

1. Saltiel, A. R. & Kahn, C. R. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **414**, 799–806 (2001).
2. Bryant, N. J., Govers, R. & James, D. E. Regulated transport of the glucose transporter GLUT4. *Nature Rev. Mol. Cell Biol.* **3**, 267–277 (2002).
3. Chiang, S. H. *et al.* Insulin-stimulated GLUT4 translocation requires the CAP-dependent activation of TC10. *Nature* **410**, 944–948 (2001).
4. Liu, J., Kimura, A., Baumann, C. A. & Saltiel, A. R. APS facilitates c-Cbl tyrosine phosphorylation and GLUT4 translocation in response to insulin in 3T3-L1 adipocytes. *Mol. Cell Biol.* **22**, 3599–3609 (2002).
5. Neudauer, C. L., Joberty, G., Tatsis, N. & Macara, I. G. Distinct cellular effects and interactions of the Rho-family GTPase TC10. *Curr. Biol.* **8**, 1151–1160 (1998).
6. TerBush, D. R., Maurice, T., Roth, D. & Novick, P. The exocyst is a multiprotein complex required for exocytosis in *Saccharomyces cerevisiae*. *EMBO J.* **15**, 6483–6494 (1996).
7. Kee, Y. *et al.* Subunit structure of the mammalian exocyst complex. *Proc. Natl Acad. Sci. USA* **94**, 14438–14443 (1997).
8. Symons, M. Rho family GTPases: the cytoskeleton and beyond. *Trends Biochem. Sci.* **21**, 178–181 (1996).
9. Tapon, N. & Hall, A. Rho, Rac, and Cdc42 GTPases regulate the organization of the actin cytoskeleton. *Curr. Opin. Cell Biol.* **9**, 86–92 (1997).
10. Van Aelst, L. & D'Souza-Schorey, C. Rho GTPases and signaling networks. *Genes Dev.* **11**, 2295–2322 (1997).
11. Chiang, S. H., Hou, J. C., Hwang, J., Pessin, J. E. & Saltiel, A. R. Cloning and functional characterization of related TC10 isoforms, a subfamily of Rho proteins involved in insulin-stimulated glucose transport. *J. Biol. Chem.* **277**, 13067–13073 (2002).
12. Hsu, S. C., Hazuka, C. D., Foletti, D. L. & Scheller, R. H. Targeting vesicles to specific sites on the plasma membrane: the role of the sec6/8 complex. *Trends Cell Biol.* **9**, 150–153 (1999).
13. Guo, W., Sacher, M., Barrowman, J., Ferro-Novick, S. & Novick, P. Protein complexes in transport vesicle targeting. *Trends Cell Biol.* **10**, 251–255 (2000).
14. Lipschutz, J. H. & Mostov, K. E. Exocytosis: the many masters of the exocyst. *Curr. Biol.* **12**, R212–R214 (2002).
15. Ting, A. E. *et al.* rSec6 and rSec8, mammalian homologs of yeast proteins essential for secretion. *Proc. Natl Acad. Sci. USA* **92**, 9613–9617 (1995).
16. Yeaman, C., Grindstaff, K. K., Wright, J. R. & Nelson, W. J. Sec6/8 complexes on trans-Golgi network and plasma membrane regulate late stages of exocytosis in mammalian cells. *J. Cell Biol.* **155**, 593–604 (2001).
17. Kimura, A., Baumann, C. A., Chiang, S. H. & Saltiel, A. R. The sorbin homology domain: a motif for the targeting of proteins to lipid rafts. *Proc. Natl Acad. Sci. USA* **98**, 9098–9103 (2001).
18. Dawson, K., Aviles-Hernandez, A., Cushman, S. W. & Malide, D. Insulin-regulated trafficking of dual-labelled glucose transporter 4 in primary rat adipose cells. *Biochem. Biophys. Res. Commun.* **287**, 445–454 (2001).

19. Novick, P. & Guo, W. Ras family therapy: Rab, Rho and Ral talk to the exocyst. *Trends Cell Biol.* **12**, 247–249 (2002).
20. Cheatham, B. GLUT4 and company: SNAREing roles in insulin-regulated glucose uptake. *Trends Endocrinol. Metab.* **11**, 356–361 (2000).
21. Chamberlain, L. H. & Gould, G. W. The v- and t-SNARE proteins that mediate Glut4 vesicle fusion are localised in detergent-insoluble lipid rafts present on distinct intracellular membranes. *J. Biol. Chem.* **277**, 49750–49754 (2002).
22. Sherman, L. A., Hirshman, M. F., Cormont, M., Le Marchand-Brustel, Y. & Goodyear, L. J. Differential effects of insulin and exercise on Rab4 distribution in rat skeletal muscle. *Endocrinology* **137**, 266–273 (1996).
23. Cormont, M. *et al.* Potential role of Rab4 in the regulation of subcellular localization of Glut4 in adipocytes. *Mol. Cell. Biol.* **16**, 6879–6886 (1996).
24. Shisheva, A. & Czech, M. P. Association of cytosolic Rab4 with GDI isoforms in insulin-sensitive 3T3-L1 adipocytes. *Biochemistry* **36**, 6564–6570 (1997).
25. Millar, C. A., Shewan, A., Hickson, G. R., James, D. E. & Gould, G. W. Differential regulation of secretory compartments containing the insulin-responsive glucose transporter 4 in 3T3-L1 adipocytes. *Mol. Biol. Cell* **10**, 3675–3688 (1999).
26. Chang, L., Adams, R. D. & Saltiel, A. R. The TC10-interacting protein CIP4/2 is required for insulin-stimulated Glut4 translocation in 3T3L1 adipocytes. *Proc. Natl Acad. Sci. USA* **99**, 12835–12840 (2002).
27. Lin, D. *et al.* A mammalian PAR-3–PAR-6 complex implicated in Cdc42/Rac1 and aPKC signalling and cell polarity. *Nature Cell Biol.* **2**, 540–547 (2000).
28. Thurmond, D. C. *et al.* Regulation of insulin-stimulated GLUT4 translocation by Munc18c in 3T3L1 adipocytes. *J. Biol. Chem.* **273**, 33876–33883 (1998).
29. Hwang, J. B. & Frost, S. C. Effect of alternative glycosylation on insulin receptor processing. *J. Biol. Chem.* **274**, 22813–22820 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements This work was supported by grants from the National Institutes of Health. We thank T.-H. Chun for helpful discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.R.S. (e-mail: saltiel@umich.edu).

Noise in eukaryotic gene expression

William J. Blake*, Mads Kærn*, Charles R. Cantor & J. J. Collins

Center for BioDynamics, Center for Advanced Biotechnology, Bioinformatics Program, and Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA

* These authors contributed equally to this work

Transcription in eukaryotic cells has been described as quantal¹, with pulses of messenger RNA produced in a probabilistic manner^{2,3}. This description reflects the inherently stochastic nature^{4–9} of gene expression, known to be a major factor in the heterogeneous response of individual cells within a clonal population to an inducing stimulus^{10–16}. Here we show in *Saccharomyces cerevisiae* that stochasticity (noise) arising from transcription contributes significantly to the level of heterogeneity within a eukaryotic clonal population, in contrast to observations in prokaryotes¹⁵, and that such noise can be modulated at the translational level. We use a stochastic model of transcription initiation specific to eukaryotes to show that pulsatile mRNA production, through reinitiation, is crucial for the dependence of noise on transcriptional efficiency, highlighting a key difference between eukaryotic and prokaryotic sources of noise. Furthermore, we explore the propagation of noise in a gene cascade network and demonstrate experimentally that increased noise in the transcription of a regulatory protein leads to increased cell–cell variability in the target gene output, resulting in prolonged bistable expression states. This result has implications for the role of noise in phenotypic variation and cellular differentiation.

To explore the effects of transcriptional variation and control on the level of noise in eukaryotic gene expression, we used both native and artificial modes of transcriptional regulation in the yeast *GAL1* promoter (Fig. 1a). In its natural context, the *GAL1* promoter is activated in response to galactose (in the absence of preferentially metabolized glucose) through an upstream activation sequence (UAS_G) composed of multiple binding sites for the transcriptional activator Gal4p. Like many eukaryotic activators¹⁷, Gal4p acts by recruiting protein complexes involved in chromatin remodelling and the ordered assembly of a transcription preinitiation complex^{18,19}. Because Gal4p is a galactose-dependent transcriptional activator, activation-based expression from the *GAL1* promoter is effectively modulated with galactose. As a second mode of transcriptional control, distinct from the native complexity of the yeast galactose-utilization pathway, we constructed an artificial, Tet-responsive *GAL1* promoter (P_{GAL1*}) by inserting tandem *tet* operators (2×*tetO*₂) downstream of the *GAL1* TATA box. In contrast to Gal4p-mediated activation, bound Tet repressor (TetR) might act by sterically hindering the assembly of the transcriptional machinery, effectively repressing expression from P_{GAL1*}. TetR-mediated repression can be relieved by the addition of the chemical inducer anhydrotetracycline (ATc), which binds directly to TetR. Constitutive expression of TetR therefore allows rheostat-like control of P_{GAL1*} transcriptional efficiency through the use of varying levels of ATc. The gene encoding the yeast-enhanced green fluorescent protein (yEGFP) was expressed from P_{GAL1*} as a quantifiable reporter, and fluorescence histograms were obtained from flow cytometric measurement of similarly sized cells containing chromosomally integrated, single genetic copies of each construct.

Transcription from P_{GAL1*} is modulated over a broad dynamic range by both native and artificial modes of regulation (Fig. 1b), allowing a direct comparison between galactose- and ATc-mediated

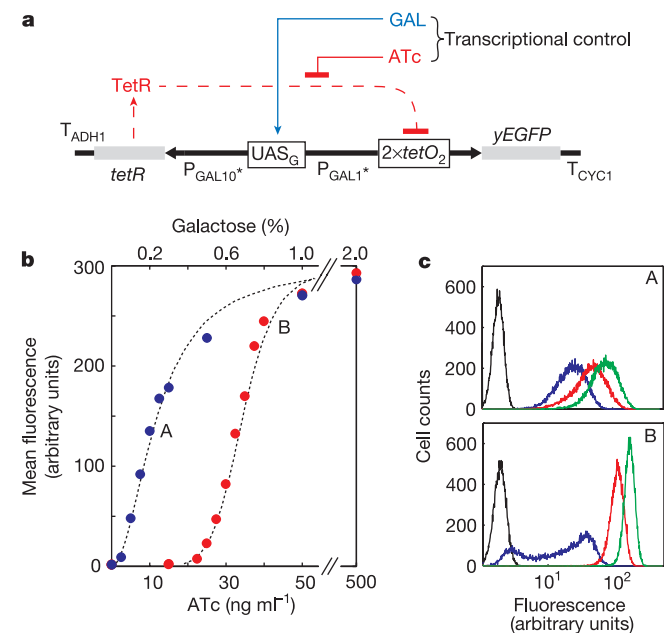


Figure 1 Transcriptional control of P_{GAL1*}. **a**, TetR, expressed from P_{GAL10*}, represses expression of yEGFP. Anhydrotetracycline (ATc) and galactose (GAL) are required to induce yEGFP expression. Transcription terminators (T_{ADH1}, T_{CYC1}) are indicated. **b**, Dose–response curve of P_{GAL1*} expressing yEGFP to ATc at full galactose induction (2%; red points), and to galactose at full ATc induction (500 ng ml^{−1}; blue points). Broken lines were obtained from stochastic simulations (Box 1). **c**, Transient response of cells marked A and B in **b**, induced with 0.2% galactose and 40 ng ml^{−1} ATc, respectively. Histograms correspond to preinduction (black), and 150 min (blue), 290 min (red) and 440 min (green) after induction.

induction of yEGFP expression from a single promoter. Each point of the dose–response curves shown in Fig. 1b represents the averaged response of an isogenic population to a particular level of induction with either ATc or galactose. Population averages were obtained from fluorescence histograms of individual cell measurements after more than 45 h of growth at a particular inducer level. Distinct modes of transcriptional control of P_{GAL1^+} yield strikingly different transient responses, with native galactose-mediated induction resulting in a graded shift in the fluorescence distribution and ATc-mediated induction causing a bimodal response (Fig. 1c). It is likely that these responses reflect the differential effects of activator and repressor proteins on the promoter itself^{20,21}. Steady-state fluorescence histograms for populations maintained at similar mean fluorescence values with either ATc- or galactose-mediated

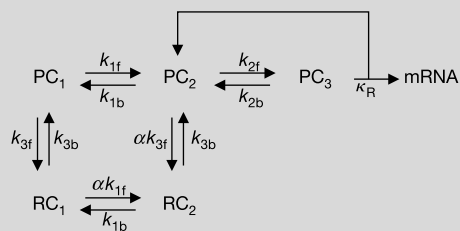
induction show considerable variation, with ATc-mediated induction causing a more heterogeneous response (histograms not shown; see Fig. 2a). We quantify this cell–cell variability in yEGFP in terms of population variance divided by population mean and refer to this value as the noise strength¹⁵.

Noise strength, measured over a wide range of P_{GAL1^+} transcriptional efficiencies, shows a striking response marked by low noise at low transcriptional efficiency, an increase at transcriptional efficiencies of 20–40%, and a gradual decrease to a low-noise state at full induction (Fig. 2a). To control for any TetR effects during galactose-mediated induction, noise strength was also measured from P_{GAL1^+} in the absence of TetR (under similar conditions of varying galactose), and was found to exhibit an identical behaviour to that of P_{GAL1^+} with full TetR induction (not shown).

Box 1

Transcriptional noise in eukaryotic gene expression

We developed a stochastic model of gene expression that incorporates several features specific to eukaryotic transcription, including sequential assembly of the core transcription apparatus, slow chromatin remodelling, rate-limiting binding of the TATA-box-binding protein (TBP) and pulsatile mRNA production due to reinitiation. The model assumes random transitions between different promoter states in transcription initiation as represented by the scheme:



The states PC_1 , PC_2 and PC_3 represent, respectively, the inactive (or silenced) promoter, an intermediate complex with TBP and various transcription factors bound, and the preinitiation complex where all components required for transcription are assembled on the promoter. Reinitiation is modelled as a return to the intermediate complex PC_2 (and subsequent transition back to PC_3) after transcription initiation has occurred from PC_3 . The states RC_1 and RC_2 represent different repressed promoter configurations. The factor α characterizes the masking of unoccupied DNA-binding sites in the repressed complex RC_1 and the intermediate complex PC_2 . Transcript elongation and translation are modelled as single-step processes with stochastic decay of mRNA and protein (see Supplementary Information).

Figure 3a shows that the model can reproduce the correlation between noise strength and transcriptional efficiency for both ATc- and galactose-mediated induction using the same parameter set (see Supplementary Information) that gives quantitative agreement with the dose–response curves (Fig. 1b). With these parameter values, the model predicts linear correlations between noise strength and translational efficiency (Fig. 3b) that are in excellent agreement with the experimental observations (Fig. 2b). The model cannot reproduce the experimental results if the probability of producing a transcript is constant (Fig. 3a, b).

Our experiments involve a chromosomally carried repressor gene, to minimize repressor fluctuations. It was previously shown in bacteria¹⁶ that substantially increased fluctuations in a controlling repressor, introduced by carrying the repressor gene on a plasmid, can result in increased levels of noise at intermediate levels of expression. As shown in Fig. 4b, the level of noise in *GFPmut3b* expression from P_{GAL10^+} ranges between about 1 and 3 arbitrary units (AU) and is about 1.3 AU at full galactose induction. Given that *GFPmut3b* and *tetR* have similar

codon adaptation indices (0.183 and 0.115, respectively), the level of noise in *GFPmut3b* expression can, when assuming similar mRNA decay rates, be used as an indicator of *tetR* gene expression noise. By using a value of 4 AU for *tetR* gene expression noise in the simulations, we overestimate repressor fluctuations and show that, in the absence of transcriptional reinitiation, these fluctuations alone (broken lines in Fig. 3a) cannot account for the noise levels observed in our experiments (Fig. 2a).

Eukaryotic activators are known to act on various steps in the transcription process²⁵, including (1) promoter activation, through remodelling of chromatin structure and/or recruitment of TBP; (2) stabilization of an intermediate, TBP-containing complex; and (3) recruitment of RNA polymerase II (RNAP). In the model, these steps correspond to an increase in k_{1f} (step 1), a decrease in k_{1b} (step 2) and an increase in k_{2f} (step 3), respectively. For example, in Fig. 3a the effect of increasing galactose (mediated through the Gal4p activator) is assumed to increase k_{1f} and decrease k_{1b} (see Supplementary Information). In Fig. 3c we explore the potential role of activators in modulating gene expression noise by varying the rates k_{1f} , k_{1b} or k_{2f} in the model.

In the model, the transcription level (the average promoter occupancy by RNAP) and the magnitude of gene expression noise are both sensitive to the extent of reinitiation, ξ , defined as the average number of mRNA transcripts produced before promoter deactivation (transition from PC_2 to PC_1):

$$\xi = \kappa_R k_{2f} / [k_{1b}(k_{2b} + \kappa_R)].$$

A promoter that has an unstable intermediate complex (high k_{1b}) that is ineffective in recruiting RNAP (low k_{2f}) lacks the ability to promote high reinitiation, causing relatively low levels of transcriptional noise. Such a promoter is also limited to low levels of transcription, even if k_{1f} could become infinitely large in the presence of an activator (Fig. 3c). Promoter strength can be substantially increased by increasing the extent of reinitiation through, for example, more efficient recruitment of RNAP (high k_{2f}). However, this causes a marked increase in noise strength at low levels of expression (Fig. 3c) owing to an average of ξ transcripts being produced in pulses at irregular intervals rather than at a low steady rate. If an activator is directly involved in the stabilization of the intermediate complex (decreasing k_{1b}) and/or the rate of RNAP recruitment (increasing k_{2f}), promoter strength can be increased markedly without affecting noise at low levels of transcription. As shown in Fig. 3c, by decreasing k_{1b} or increasing k_{2f} , an activator can cause transcriptional noise to pass through a maximum when the promoter is about 50% activated with low transcriptional noise at both low and high levels of transcription. By acting on multiple steps in the transcription initiation process, transcriptional activators can be used to modulate gene expression noise and ensure tight basal expression, without compromising promoter strength or dynamic range.

The observed difference between noise signatures for ATc- and galactose-mediated induction might therefore reflect the different molecular mechanisms by which galactose and ATc induce promoter activation, as shown by transient graded and binary responses, respectively (Fig. 1c). To determine whether the qualitative characteristics of the observed noise signatures are specific to P_{GAL1^*} , we engineered a second Tet-responsive promoter by inserting $2\times tetO_2$ into the yeast *ADHI* promoter in the same location, relative to the TATA box, as in P_{GAL1^*} . The engineered *ADHI* promoter (P_{ADHI^*}) is 10-fold weaker than P_{GAL1^*} at full ATc induction, but it shows a non-monotonic transcriptional noise signature similar to those observed for P_{GAL1^*} (Fig. 2a, inset). P_{ADHI^*} also exhibits transient bimodality when induced with ATc (not shown), offering further evidence that the mode of transcriptional control can have a marked effect on the resulting population response.

Recent bacterial studies^{15,16} have explored the contribution of transcription to the level of noise in prokaryotic gene expression. It was demonstrated experimentally, with a chromosomally integrated LacI-repressible promoter, that the level of transcription (modulated with isopropyl β -D-thiogalactoside inducer) had little effect on the level of noise¹⁵, demonstrated by a weak positive correlation between transcriptional efficiency and noise strength. The non-monotonic responses to varying levels of transcription exhibited by eukaryotic cells in the present study (Fig. 2a) contrast sharply with what was observed in a prokaryote¹⁵, in which a similar noise measure and single-copy constructs make comparison amenable. One of the many factors that differ between prokaryotic and eukaryotic gene expression is the rate-limiting step of transcriptional initiation, which, for eukaryotic organisms, involves the binding of the TATA-box-binding protein (TBP) to the promoter region^{22,23}. Highly expressed genes are probably associated with promoters that employ the process of transcriptional reinitiation²⁴, whereby a stabilized complex (including TBP) remains on the promoter for multiple rounds of transcription. By obviating the need to reassemble the preinitiation complex *de novo*, transcriptional reinitiation allows the faster production of a large amount of transcript.

The stochastic model presented in Box 1, and described in detail in Supplementary Information, was used to explore specific mechanisms of eukaryotic transcription apparatus assembly, as well as the process of reinitiation, which may contribute to the shape and magnitude of the observed noise signatures. The model incorporates random transitions between various states of promoter occupancy by activators, repressors and general transcription factors in the process of forming a functional preinitiation complex before

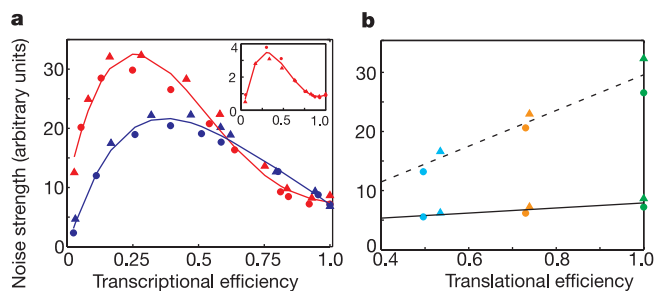


Figure 2 Effect of transcriptional and translational efficiency on noise strength. Data points represent measurements taken at >30 h (circles) and >45 h (triangles) of growth. **a**, Transcriptional efficiency from P_{GAL1^*} is varied with either ATc (red points) or galactose (blue points). The inset shows noise strength as a function of transcriptional efficiency (varied by ATc induction) for P_{ADHI^*} . Fourth-order polynomial fits are included for clarity. **b**, Variants *yEGFP* (green points), *yEGFPm1* (orange points) and *GFPmut3b* (light blue points) differ only in codon content (CAI = 0.596, 0.384 and 0.183, respectively). Linear fits to data obtained at full transcriptional induction (solid line, slope = 4.3), and at ~28% (30 ng ml⁻¹ ATc) transcriptional induction (broken line, slope = 30.2) are shown.

transcriptional initiation and reinitiation. Importantly, simulations of the model (Fig. 3a) demonstrate that pulsatile mRNA production (resulting from transcriptional reinitiation) is required to reproduce the experimental observations, providing further support for the quantal view of eukaryotic transcription¹ (see Box 1). An additional important factor in the control of eukaryotic gene expression is the role of transcriptional activators, which, through a variety of mechanisms²⁵, can affect the rates of transition between various states of promoter occupancy and accessibility. We use the model to explore these rates, as potential targets of activator action, and show how activators can modulate gene expression noise (see Box 1).

Translational efficiency, or the number of protein molecules produced from a single mRNA transcript, has been proposed^{15,7}, and recently shown¹⁵, to be an important variable in determining the level of noise in prokaryotic gene expression. To determine whether translational efficiency contributes to the level of noise in eukaryotic gene expression, we modified the construct shown in Fig. 1a by replacing the *yEGFP* gene with two codon variants, *GFPmut3b* (the precursor to *yEGFP*) and *yEGFPm1* (a hybrid gene containing sequences from *yEGFP* and *GFPmut3b*). Codon usage, as measured by the codon adaptation index (CAI)²⁶, is a reliable indicator of translational efficiency in yeast and is easily modified without disrupting the amino-acid sequence or functionality of the protein. Relative to *yEGFP* at full (100%) induction from P_{GAL1^*} , *GFPmut3b* and *yEGFPm1* showed low (~50%) and intermediate (~75%) expression, respectively, which correlate well with the calculated CAI values for each of these variants (Fig. 2b). As shown in Fig. 2b, an increase in the translational efficiency of the reporter causes a slight increase in the noise strength at full induction of P_{GAL1^*} . However, increased translational efficiency

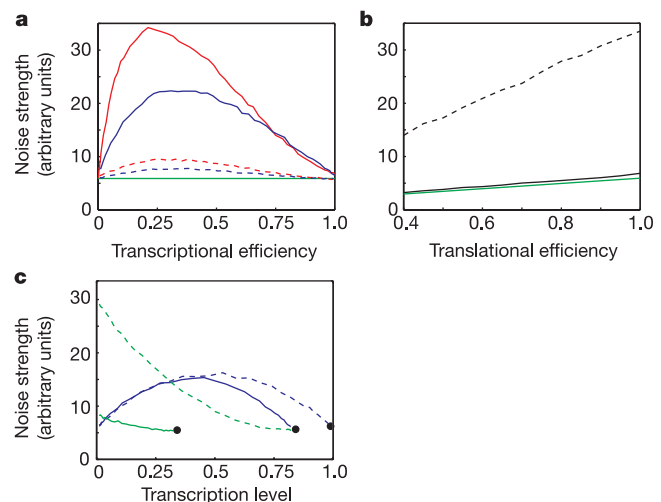


Figure 3 Simulations of eukaryotic gene expression noise. **a**, Correlation between noise strength and transcriptional efficiency varied with ATc (red curve), galactose (blue curve), and for constant probability of producing a transcript (green line). Simulations in the absence of transcriptional reinitiation are shown (broken curves). **b**, Correlation between noise strength and translational efficiency for full (solid line, slope = 5.7) and ~28% (broken line, slope = 33) transcriptional induction. The green line (slope = 4.9) is obtained when the probability of producing a transcript is constant. **c**, Correlation between level of transcription and noise at different model parameter values. The green curves show variation in the level of transcription through k_{1f} for low (solid curve, $k_{1b} = 1$, $k_{2f} = 1$, $\xi = 0.5$) and high (broken curve, $k_{1b} = 1$, $k_{2f} = 10$, $\xi = 5$) extents of reinitiation. The blue curves show variation in the level of transcription through k_{1b} for $k_{1f} = 0.1$, $k_{2f} = 10$ (solid curve); and through k_{2f} for $k_{1f} = 0.1$, $k_{1b} = 1$ (dotted curve). The level of transcription is reported relative to the highest level of mRNA theoretically possible. Repressor is absent ($k_{3f} = 0$) and RNA polymerase II binds tightly to the promoter ($k_{2b} = 1$).

has a substantial effect when coupled to a noisy transcriptional state ($\sim 28\%$ induction), amplifying transcriptional noise considerably. This provides additional evidence that the level of gene expression noise in eukaryotic cells is strongly influenced by transcription, in contrast to observations in prokaryotic cells¹⁵. Strikingly similar results are produced with the stochastic model (Fig. 3b). The results in Fig. 2b show that translational efficiency (in particular, codon usage) can be used to regulate the level of noise in the expression of a eukaryotic gene; these findings indicate that noise strength is an important factor to be considered when studying the role of codon bias in gene expression.

There has been speculation about the effects of a noisy regulatory input on downstream gene expression^{7,9,27}, and it has been proposed that important regulatory genes have evolved to minimize noise¹⁵. To determine the extent to which fluctuations in the expression of a regulatory gene can affect cell–cell variability in the downstream target, we constructed a simple regulatory cascade (Fig. 4a) in which noise in the expression of the regulator (TetR) can be modulated independently of the regulated gene (*yEGFP*). Independent noise control is feasible because of the differential effect of galactose on the *GAL10* and *ADH1* promoters. Specifically, P_{ADH1^+} expression is not regulated directly by galactose and the level of noise in transcription from P_{ADH1^+} does not change significantly when galactose is varied (Fig. 4b). The level of noise in TetR (expressed

from P_{GAL10^+}) can therefore be varied with galactose independently of noise in the expression from P_{ADH1^+} . We set the tunable noise in TetR expression to either high (0.15% galactose) or low (2% galactose) (see Fig. 4b), and used ATc to control the regulatory effect of TetR on P_{ADH1^+} transcriptional efficiency. Figure 4c shows that increased noise in TetR expression has a marked effect on the level of noise in ATc-regulated *yEGFP* expression, causing a substantial increase in target gene fluctuations over a broad range of transcriptional efficiencies.

As described above, ATc-induced expression of *yEGFP* from P_{ADH1^+} results in a transient bimodal response (similar to that of P_{GAL1^+} shown in Fig. 1c), which becomes unimodal when expression reaches a steady state. By directly tuning the level of noise in the TetR regulatory protein, independently of other factors involved in expression from P_{ADH1^+} , we find that we can directly affect *yEGFP* expression state stability, as reflected in the maintenance of a bimodal response. Specifically, a low level of noise in TetR expression (noise strength ~ 1.3 arbitrary units (AU)) causes *yEGFP* expression to exhibit a single, unimodal fluorescence distribution (Fig. 4d), whereas when the TetR regulatory protein is tuned to a high noise level (noise strength ~ 2.7 AU), we observe that *yEGFP* exhibits prolonged bistable expression states that persist for the length of the experiment (Fig. 4e). These results show that the downstream effects of increased cell–cell variability in a regulatory protein can have profound phenotypic consequences, drastically affecting the stability of gene expression states. It was recently shown that increased noise in the expression of a tumour-suppressor gene can lead to altered cell phenotypes characterized by distinct morphological changes²⁸. The cascading noise experiments described above demonstrate a plausible mechanism for such heritable changes in target gene expression, whereby increased levels of noise in the expression of a regulatory protein can cause a population of isogenic cells to exhibit prolonged bistable expression states.

We have shown how transcriptional efficiency and the mechanism of transcriptional control relate to noise strength in certain eukaryotic promoters, providing an important distinction between sources of noise in prokaryotic and eukaryotic cells. We have also demonstrated a clear link between codon usage and gene expression noise. Moreover, we have shown that increased noise in a regulatory protein can have profound effects on bimodal responses that can be critical for cellular differentiation and the maintenance of phenotypic variation. Together these findings contribute to a greater understanding of the origins of cell–cell variability and the consequences of such variability in the expression of a regulatory protein on cell phenotype. Our stochastic model of gene induction proposes specific mechanisms by which slow transitions between various states of transcription apparatus assembly and the process of reinitiation, and also the action of transcriptional activators, might modulate noise in gene expression. In addition to the proposed role of activators, the level of noise might involve other factors, such as the sequence of the promoter itself. The stability of the complex between transcription factor IIA, TBP and promoter is dependent on the sequence of the TATA box²⁹, and promoters that direct reinitiation have consensus TATA elements³⁰. Variation in TATA box sequence and the use of TATA-less promoters might be additional mechanisms of noise control, as a high extent of reinitiation (caused by a stable intermediate promoter complex) can lead to high noise (see Box 1). All of these mechanisms are known to be involved in eukaryotic gene expression and might work synergistically to control the level of noise. □

Methods

Strains, media and growth

All yeast strains were created by targeted chromosomal integration of shuttle vector constructs at either the *GAL1-10* or the *ADH1* locus of *S. cerevisiae* strain YPH500 (α , *ura3-52*, *lys2-801*, *ade2-101*, *trp1Δ63*, *his3Δ200*, *leu2Δ1*) (Stratagene) or derivatives,

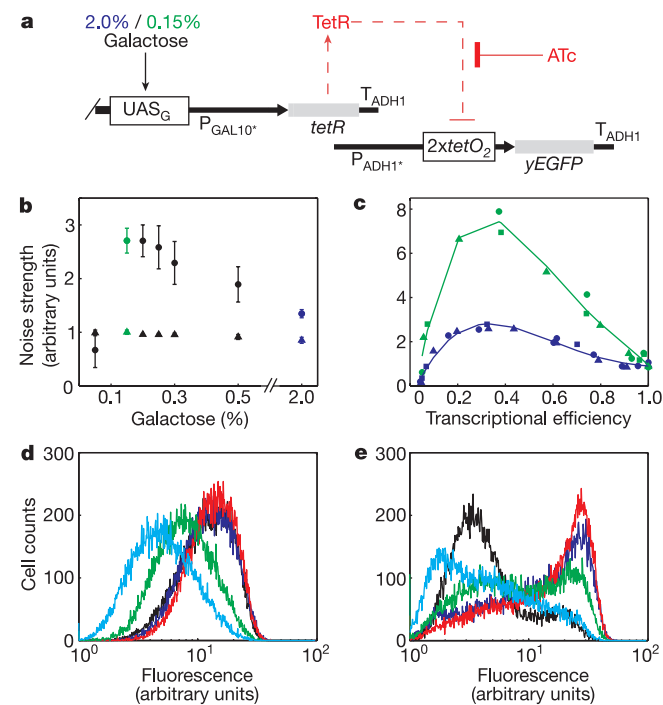


Figure 4 Cascading noise in a gene network and effect on cellular response. **a**, Diagram of simple cascading gene network with regulatory input (TetR) and regulated output (*yEGFP*). **b**, Noise strength as a function of galactose for P_{GAL10^+} expressing *GFPmut3b* (circles) and P_{ADH1^+} expressing *yEGFP* (triangles). Conditions of high (0.15% galactose; green points) and low (2% galactose; blue points) input noise are indicated. *GFPmut3b* (CAI = 0.183) is used to estimate the effect of galactose on noise in the expression of *tetR* (CAI = 0.115). Error bars are 95% confidence intervals for three independent measurements at >15 h, >30 h and >45 h induction with galactose. **c**, Noise in *yEGFP* expression as a function of transcriptional efficiency at high (blue points) and low (green points) levels of input noise. Data points represent day 4 (circles), day 5 (triangles) and day 7 (squares) measurements. Fourth-order polynomial fits are included for clarity. **d, e**, Histograms represent *yEGFP* fluorescence measurements taken after growth for 1 day (black), 2 days (dark blue), 3 days (red), 4 days (green) and 5 days (light blue). Low (d) and high (e) levels of TetR input noise cause different phenotypic responses.

and confirmed by polymerase chain reaction (PCR) analysis. Cultures were grown in synthetic drop-out (SD) medium supplemented for selection of correct integrants and containing either 2% galactose with various amounts of ATc inducer, or 500 ng ml⁻¹ ATc and 2% raffinose with various amounts of galactose as an inducer. For cascading noise experiments, all cultures contained 2% raffinose. Exponentially growing yeast cells were diluted 1:30 to 1:50 into SD medium containing the appropriate inducer concentrations and grown at 30 °C for 14–19 h to an OD₆₀₀ of 0.6 ± 0.3 for flow cytometric analysis. Inducer concentrations were maintained in subsequent dilutions, and cells were assayed every 14–19 h for 3–7 d. All cultures containing ATc were protected from light during growth. All cloning steps were performed in *Escherichia coli* XL10-Gold (Stratagene).

Artificial switch construction

All plasmid backbones were derived from pRS403 or pRS404 shuttle vectors (Stratagene). The *GAL1-10* promoter region was obtained from pESC-Leu (Stratagene) and the *ADHI* promoter region from pAD4Δ (a gift from T. Gilmore, Boston University). Promoters were modified by PCR-directed insertion of bacterial operator sites, 2XtetO₂ or *lacO*, downstream of the *GAL1*, *ADHI* or *GAL10* TATA boxes to create P_{GAL1}, P_{ADHI} and P_{GAL10}, respectively (only P_{GAL10} contains *lacO*). Genes encoding TetR, yEGFP and GFPmut3b were amplified by PCR from pcDNA6/TR (Invitrogen), pEGFP3 (a gift from B. Cormack, Johns Hopkins University) and pGFP(LVA) (Clontech), respectively. The yEGFPm1 gene was created by inserting the *MscI*–*HpaI* fragment from GFPmut3b into yEGFP. All *gfp* variants were sequenced with twofold coverage to ensure identical amino-acid sequences. CAI, a measure of codon bias, was calculated with CodonW (J. Peden, University of Nottingham, UK; <http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html>).

Cytometry and data analysis

Expression data were obtained with a Becton-Dickinson FACSCalibur flow cytometer with a 488-nm argon excitation laser and a 515–545-nm emission filter (FL1). Cultures in mid-exponential-phase growth were pelleted and resuspended in filtered PBS before assay. Cell samples were assayed at a low flow rate until 30,000 cells had been collected within a small forward scatter and side scatter gate to minimize fluorescence variation due to cell size. Flow cytometry standard list-mode files were converted to ASCII format with MFI (E. Martz, University of Massachusetts, Amherst; <http://www.umass.edu/microbio/mfi>) and analysed by using Matlab (The MathWorks, Inc., Natick, Massachusetts). Transcriptional efficiency was defined relative to the fully induced (500 ng ml⁻¹ ATc, 2% galactose) case for each measurement day, whereas translational efficiency was defined relative to yEGFP expression for each measurement day.

Simulations

Simulations of the reaction network described in Box 1 were performed with a stochastic algorithm as described in detail in the Supplementary Information. Mean values and variances were calculated from histograms obtained from 10⁴ independent realizations of length 10³ dimensionless time units (18-fold the protein half-life). For computational efficiency, fluorescence was assumed to equal the number of reporter protein molecules. With the exception of TetR and galactose, nuclear concentrations of transcription factors and cofactors were assumed to be constant and absorbed into the pseudo-first-order association (*k_{if}*) and dissociation (*k_{ib}*) rate constants. Cell–cell variability in rates of transcription elongation, translation and decay of mRNA and protein cause an increase in the noise strength at increased transcriptional efficiency (not shown), and are assumed to be low. The level of galactose in each individual cell was obtained by drawing a random value from a normal distribution with a standard deviation of 10% of the mean. Fluctuations in TetR protein were modelled using a two-step model of transcription and translation¹⁵. Broken lines in Fig. 3a were obtained with a model incorporating fast promoter state transitions with fluctuations in TetR (red) and cell–cell variability in galactose (blue), respectively (see Supplementary Information). The green lines in Fig. 3a, b were obtained from the same model in the absence of galactose and TetR cell–cell variability, and hence show a time-independent probability of producing a transcript. In Fig. 3b, translational efficiency was varied by changing the rate constant *k_p* associated with translation (see Supplementary Information). The green line in Fig. 3b was obtained with 30 ng ml⁻¹ ATc and 2% galactose. A coinciding curve is obtained with 500 ng ml⁻¹ ATc and 2% galactose. In Fig. 3c, *k₁₆*, *k_{1b}* and *k_{2f}* were varied between zero and infinity.

Received 20 December 2002; accepted 7 March 2003; doi:10.1038/nature01546.

- Hume, D. A. Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. *Blood* **96**, 2323–2328 (2000).
- Ross, I. L., Browne, C. M. & Hume, D. A. Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol. Cell Biol.* **72**, 177–185 (1994).
- Walters, M. C. *et al.* Enhancers increase the probability but not the level of gene expression. *Proc. Natl Acad. Sci. USA* **92**, 7125–7129 (1995).

- Ko, M. S. H. A stochastic model for gene induction. *J. Theor. Biol.* **153**, 181–194 (1991).
- McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl Acad. Sci. USA* **94**, 814–819 (1997).
- Hasty, J., Pradines, J., Dolnik, M. & Collins, J. J. Noise-based switches and amplifiers for gene expression. *Proc. Natl Acad. Sci. USA* **97**, 2075–2080 (2000).
- Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl Acad. Sci. USA* **98**, 8614–8619 (2001).
- Kepler, T. B. & Elston, T. C. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* **81**, 3116–3136 (2001).
- Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl Acad. Sci. USA* **99**, 12795–12800 (2002).
- Novick, A. & Weiner, M. Enzyme induction as an all-or-none phenomenon. *Proc. Natl Acad. Sci. USA* **43**, 553–566 (1957).
- Ko, M. S. H., Nakachi, H. & Takahashi, N. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *EMBO J.* **9**, 2835–2842 (1990).
- Fiering, S. *et al.* Single cell assay of a transcription factor reveals a threshold in transcription activated by signals emanating from the T-cell antigen receptor. *Genes Dev.* **4**, 1823–1834 (1990).
- van Roon, M. A., Aten, J. A., van Oven, C. H., Charles, R. & Lamers, W. H. The initiation of hepatocyte-specific gene expression within embryonic hepatocytes is a stochastic event. *Dev. Biol.* **136**, 508–516 (1989).
- Becskei, A., S  raphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20**, 2528–2535 (2001).
- Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31**, 69–73 (2002).
- Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
- Plashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–577 (1997).
- Bh  m  k, S. R. & Green, M. R. SAGA is an essential *in vivo* target of the yeast acidic activator Gal4p. *Genes Dev.* **15**, 1935–1945 (2001).
- Larschan, E. & Winston, F. The *S. cerevisiae* SAGA complex functions *in vivo* as a coactivator for transcriptional activation by Gal4. *Genes Dev.* **15**, 1946–1956 (2001).
- Rossi, F. M. V., Kringstein, A. M., Spicher, A., Guicherit, O. M. & Blau, H. M. Transcriptional control: rheostat converted to on/off switch. *Mol. Cell* **6**, 723–728 (2000).
- Biggar, S. R. & Crabtree, G. R. Cell signaling can direct either binary or graded transcriptional responses. *EMBO J.* **20**, 3167–3176 (2001).
- Chatterjee, S. & Struhl, K. Connecting a promoter-bound protein to TBP bypasses the need for a transcriptional activation domain. *Nature* **374**, 820–822 (1995).
- Klages, N. & Strub  n, M. Stimulation of RNA polymerase II transcription initiation by recruitment of TBP *in vivo*. *Nature* **374**, 822–823 (1995).
- Struhl, K. Chromatin structure and RNA polymerase II connection: implications for transcription. *Cell* **84**, 179–182 (1996).
- Hahn, S. Activation and the role of reinitiation in the control of transcription by RNA polymerase II. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 181–188 (1998).
- Sharp, P. M. & Li, W. H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987).
- Thattai, M. & van Oudenaarden, A. Attenuation of noise in ultrasensitive signaling cascades. *Biophys. J.* **82**, 2943–2950 (2002).
- Kemkemmer, R., Schrank, S., Vogel, W., Gr  ler, H. & Kaufmann, D. Increased noise as an effect of haploinsufficiency of the tumour-suppressor gene neurofibromatosis type 1 *in vitro*. *Proc. Natl Acad. Sci. USA* **99**, 13783–13788 (2002).
- Stewart, J. J. & Stargell, L. A. The stability of the TFIIA-TBP-DNA complex is dependent on the sequence of the TATAAA element. *J. Biol. Chem.* **276**, 30078–30084 (2001).
- Yean, D. & Gralla, J. Transcription reinitiation rate: a special role for the TATA box. *Mol. Cell. Biol.* **17**, 3809–3816 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

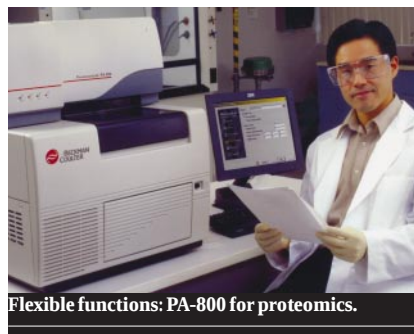
Acknowledgements We thank T. Gilmore for guidance and helpful advice on switch construction, F. Isaacs for helpful discussions and advice, and B. Cormack for the gift of pEGFP3. This work was supported by DARPA, NSF and the Danish Research Agency.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.J.C. (e-mail: jjcollins@bu.edu).

Proteins and antibodies

Advanced handling and no messing.



Flexible functions: PA-800 for proteomics.

ProteomeLab PA-800

Beckman Coulter www.beckmancoulter.com

Protein characterization, automatically

Beckman Coulter's PA-800 system offers a comprehensive range of automated protein characterization functions, including molecular weight determination, peptide mapping, isoelectric focusing, carbohydrate profiling and mechanisms for the study of protein-protein interactions in solution. The system can manipulate and resolve low levels of proteins that range from acidic to basic to membrane bound. The PA-800 is part of ProteomeLab, the company's suite of proteomic tools for cellular identification and protein isolation, fractionation, characterization and evaluation.

Thio-Link

Severn Biotech www.severnbiochem.com

Spaced out

Thio-Link affinity purification gel can be used to purify anti-peptide antibodies directly from sera. The gel has both a 10-atom and a 12-atom spacer arm to which the antigen is immobilized or bound using the thiol (SH) group of the peptide or protein



Bottle bank: Thio-Link antibody purifier.

antigen. The spacer reduces steric hindrance and results in a high binding capacity, coupling up to 6 mg of cysteine amino acid per ml of gel. Thio-Link gel is available alone or as part of a kit.

AKTApilot

Amersham Biosciences

www.amershambiosciences.com

Protein production on the bench

The AKTApilot unit is a benchtop separation system that links bench-scale process development with large-scale protein manufacturing. It is particularly suitable for carrying out process development through small-scale production on the same unit. The system offers a range of flow rates, from 4 to 800 ml per minute, and allows for multiple wavelength detection as well as pH and conductivity detection. Eight outlet valves handle fractionation. An optional Frac-950 can be accommodated, as can an additional monitor for special detection requirements. The UNICORN operating system enables the transfer of methods for both scale-up and scale-down applications.

Hand Operator

Integra Biosciences

www.integra-biosciences.com

High aspirations

Integra's Hand Operator converts almost any commercial vacuum pump into a flexible system for aspirating liquids from a wide range of sample media, including 96- and 24-well plates, culture bottles, sample tubes and vials. The device comes with a range of adapters for different applications and is compatible with most commercial pipette tips and stainless steel needles. A knob provides sensitive regulation of the applied vacuum. A combination of Integra's VacuSafe pump with the Hand Operator provides a 'one-stop' system for safe and productive aspiration of liquid samples. The VacuSafe is a compact pump station with integrated controller and waste container.

SOLO cell culture system

FiberCell Systems www.fibercellsystems.com

Hollow promise

FiberCell's SOLO hollow fibre cell culture system is a closed continuous flow bioreactor that can achieve cell densities over 10^8 ml^{-1} and produce 100 mg–2 g of protein. Various molecular weight cut-off hollow fibre cartridges can be used to maximize the selection and yield of secreted proteins. Appropriate for both adher-



Going SOLO can give high cell density.

ent and suspension cell lines, the SOLO system includes a single microprocessor-controlled pump that can produce flow rates ranging from 2.5 to 80 ml per minute, and is designed to fit easily in most CO₂ incubators. It can support all types and sizes of hollow fibre cartridges, from research-scale to 1.2 m² production-scale. Applications include production of monoclonal antibodies, recombinant proteins, adenovirus vectors, lymphocytes, cytokines and growth factors and endothelial cell cultures, as well as artificial organ research.

Purifier vertical clean benches

Labconco www.labconco.com

Protection from top to bottom

Labconco's 2- and 3-foot vertical benches are designed for plant tissue culture, media plate preparation, electronics inspection, medical device assembly and pharmacy drug preparation under ISO Class 5 conditions to prevent airborne contamination. While operat-



High living: Vertical benches can be for cultures.

ing, room air is drawn through a prefilter in the top of the cabinet to trap large particles. The air then flows through a HEPA filter and is projected vertically over the work area, preventing contaminants from entering and providing a particulate-free work environment. Indicator lights monitor differential pressure across the HEPA filter. The angled front sash tilts up for cleaning and pivots down to the operating position of 9 inches. Glare-free fluorescent lighting is located outside the work area. Purifier clean benches cannot be used in conjunction with biohazardous material, toxins or radionuclides.

Zymo-Film

Wako Chemicals www.wako-chemicals.de

MMP observation in situ

Zymo-Film was developed for the observation of the enzyme activity of matrix metalloproteinases (MMPs) in tissue specimens. MMPs are known to be an indicator of carcinogenesis and other physiological conditions. The *in situ* film is composed of a layer of special gelatine in a polyester base.

Protein precipitation filtration plate

Porvair Sciences www.porvair-sciences.com

Nobody messes with this one

Porvair Sciences' filtration plate uses a dual filter design to eliminate the mess and waste from leakage in traditional protein precipitation plates. A prefilter retains large flocculent particles, avoiding 'plugging', and a final filter traps residual protein particles. The hydrophobic/oleophobic filters retain the acetonitrile/sample protein precipitation mix for complete mixing without loss or leakage. A larger-than-normal pore size can be used for an improved flow rate. The protein precipitation filtration plate allows for processing 96 samples per batch; a 70% time saving over traditional methodologies is typical. It can be integrated into automated systems.

GeBAflex-tube

Accurate Chemical & Scientific

www.accuratechemical.com

Versatile preparation method

GeBAflex-tube is a gel dialysis and electroelution system suitable for all characteristic



GeBAflex-ible for characteristic proteins.

proteins and specifically efficient for hydrophobic protein. The manufacturer claims that protein is extracted and purified with greater efficiency and that SDS is eliminated by its detergent buffer, resulting in improved MALDI-MS analysis quality.

Apex 396 automated peptide synthesizer Advanced ChemTech

www.advancedchemtech.com

Robotic precision

Available in four different robotic configurations, the Apex 396 benchtop peptide synthesizer has dual robotic arms for rapid, pre-

cise delivery and advanced liquid handling techniques, eliminating impurities and contamination. It incorporates an indexed tabletop and users can select the best locations for reagent bottles. The Apex 396 can be configured to synthesize between 1 and 96 peptides and accommodates a variety of scales. Stainless steel lines the interior to minimize the effects of chemical corrosion, and an enclosed cabinet can function as a fume hood when connected to an exhaust source. Standard Fmoc and Boc coupling can be produce amounts of peptide ranging from 5 mg to 2 g, with on-board TFA cleavage for Fmoc chemistry. Applications include drug discovery, multiple analogues, combinatorial libraries, SAR studies, amino acid substitution scanning, epitope mapping and organic synthesis.

Q Trap

Applied Biosystems

www.appliedbiosystems.com

High-performance mass spectrometer

The Q Trap system is a linear, hybrid mass spectrometer in which two spectrometry technologies are combined—triple quadrupole and ion trap—to improve performance in phosphorylation site identification and protein identification. It produces rapid and sensitive full scans and features real-time charge state determination, with Analyst software for automated acquisition and Pro ID software for automated protein identification.

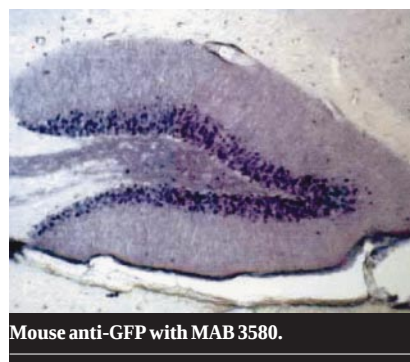
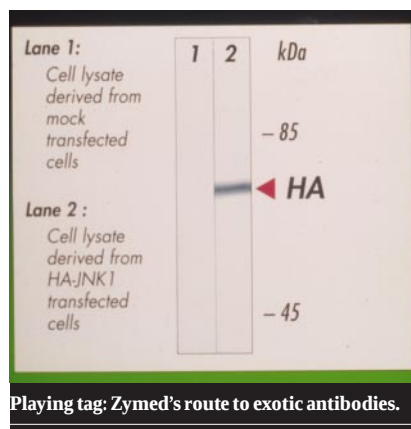
Epitope tag antibodies

Zymed

www.zymed.com

For the antibody you want

Zymed offers a full line of epitope tag antibodies for use in situations where there is no commercially available antibody against a protein of interest. Ligation of the epitope tag-encoding DNA to the protein of interest in an expression vector generates a fusion protein that can subsequently be detected with anti-epitope tag antibodies.



MAB 3580

Chemicon

www.chemicon.com

Green light for gene expression

The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is used as a fluorescent indicator for monitoring gene expression in a variety of cellular systems, including living organisms and fixed tissues. Chemicon's MAB3580 monoclonal antibody for the detection of GFP in experimental systems is reactive to both native and recombinant sources. The antibody has been affinity-purified against GFP to maximize specificity and sensitivity. Applications include immunohistochemistry, immunocytochemistry and western blot at dilutions ranging from 1:500 to 1:5,000.

Everest

Grace Vydac

www.gracevydac.com

Protein-peptide separation scales the peak

Everest high-resolution, reverse-phase, silica-based columns and media are designed for protein-peptide separations. A bonding technology is employed to maximize peak resolution of complex digests. The columns and media are said to last longer and have greater sensitivity and selectivity, with proven column-to-column and lot-to-lot reproducibility. The products can be scaled to suit applications from nano to process level.

μMACS Protein A or G MicroBeads

Miltenyi Biotec

www.miltenyibiotec.com

Pure protein in two hours

μMACS Protein A and G MicroBeads for small-scale immunoprecipitation of proteins bind to the Fc region of the specific antibody so that the immune complex is magnetically labelled and separated. Pure protein can be produced in less than two hours. The method is said to be both quick and easy, and can be seen on the company's website.

These notes are compiled in the Nature office from information provided by the manufacturers.

Material intended for publication can be sent to newproducts@nature.com.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakakamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejp.com

naturejobs

In Franklin's footsteps

In Britain, Susan Gibson is in a minority. She is a chemistry professor — a rarity in that only 5% of the country's maths, science and engineering professors are women. But Gibson, a synthetic chemist at King's College London, is on a mission to redress the balance by encouraging more women to consider a career in science.

At the end of last month, Gibson won the first Rosalind Franklin Award, a prize set up by the UK Department of Trade and Industry specifically to promote women in science. The award, worth £30,000 (US\$47,000), will allow her to start a programme to introduce students to successful women scientists from around the world.

As well as using her prize money to fund a female postdoc in her lab, Gibson plans to bring women scientists from overseas to visit labs and universities throughout Britain. Perhaps most importantly, Gibson will also arrange for her guests to talk to undergraduates and high-school students who have an interest in science. Bringing women scientists into the classroom will help young students to realize that "there are really, really exciting careers out there", Gibson says.

Gibson estimates that she has sufficient funds to bring over three women scientists each year for the next two or three years. She plans to increase the scope of the programme by attracting industrial support. That shouldn't be too difficult, as companies are beginning to realize that women are under-represented in the commercial sector too.

The success of the programme may well depend on how much support it can win from industry and whether other universities adopt it — but Gibson is optimistic that she will soon be in less of a minority.

Paul Smaglik
Naturejobs editor



Contents

REGIONS

Paris breathes life
into science

p644

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

University challenge

Paris

Spreading out: the Pierre and Marie Curie University will soon gain more space at its Jussieu campus, when Denis Diderot University relocates to the ZAC Rive Gauche.

Paris may be the 'city of light' for tourists, but the skies have been cloudy for scientists in recent years. More than a decade of government efforts to decentralize research from the Ile-de-France — the area around Paris — has resulted in neglect for the place where nearly half of the country's research is carried out.

Now politicians are realizing that the area can no longer coast along on its tradition of excellence. Thanks in part to regional funding, the Ile-de-France is building public-private partnerships and developing research in genomics and optics. The politicians now hope that jobs and venture capital will follow.

One of the biggest projects is the renovation of the Left Bank to the east of the city. The area, known as the 'ZAC Rive Gauche', is lined with warehouses, some of which will house a multidisciplinary university complex to be completed by 2005. A start-up 'incubator' and small-business centre are also planned.

The project will give France's largest scientific institute, the Pierre and Marie Curie University, space to

breathe. At present, many of its 4,000 researchers and 30,000 students share a dilapidated campus at Jussieu, in the centre of Paris, with Denis Diderot University. But the latter is uprooting to the new site, giving the Pierre and Marie Curie University more room.

Gilbert Béréziat, vice-chancellor at the Pierre and Marie Curie University, sees the opening up of the Jussieu campus as a chance to reinvent the university, with more emphasis on training people for non-traditional science careers. The campus will also be rejuvenated: some 100 research and teaching posts will become available in each of the next few years because many people will be retiring — a pattern seen throughout French public research.

Béréziat emphasizes the importance of industrial partnerships. "We have to protect some of our budget," he says, referring to a current government freeze on public laboratories. A 20-strong industrial activities unit has just been set up and the university is a partner in Agoranov, a multidisciplinary incubator that has already supported 50 start-ups in three years.

P. KITMACHER

Luring staff to Soleil

As lorries move in to prepare the site on the Saclay Plateau, south of Paris, for France's third-generation synchrotron, Soleil, the recruitment process is already well under way.

Soleil is scheduled to start operating in 2006, but researchers from public labs face a hard choice. Soleil is structured as a private company, with the French national research agency (CNRS) and the atomic

energy commission as its main shareholders. Some public-sector researchers who will move from the neighbouring LURE synchrotrons, when these close at the end of the year, protested that they would lose out in terms of career progression if they were forced to take a private contract. A compromise means they will now be able to choose to keep their public-sector status, or

take a private contract and a salary that is 15–25% higher.

Out of 350 staff, 95 researchers will be taken on gradually as the 24 beamlines become operational. With beams ranging from ultraviolet to soft and hard X-rays and a number of lines dedicated to biological research, the recruiters hope to lure staff from many disciplines. S.G.
www.synchrotron-soleil.fr

NURTURING GROWTH

New incubators popping up in Paris show that the public sector is now willing to engage in technology transfer and is being helped to do so. The city and regional councils are helping to fund incubators and locate premises for start-ups. A long-awaited new law will give tax breaks to start-ups for their first eight years. Even the pharmaceutical giant Aventis Pharma, which is reducing its own workforce, is banking on an innovation park to safeguard employment and expertise on its site (see 'Aventis alternatives', opposite).

Incubators can nurture companies in France: Hybrigenics, a successful functional-proteomics company, started life at the Pasteur Institute's incubator in Paris before the city council helped the company to find bigger premises. Belgian chief executive Donny Strosberg previously ran another start-up at France's biggest biotechnology cluster, Genopole, 30 km south of Paris, but says it was too far from the city's hub. "Access is really important," he says. "We need to be close to the laboratories we're working with."

Despite its suburban location, Genopole is holding

REGIONS

its own. Created in 1998 with massive funding from regional councils, Genopole has rapidly expanded to incorporate some 23 genomics, proteomics and bioinformatics laboratories — including Génoscope, which sequenced chromosome 14 for the Human Genome Project. One of its main laboratories, Généthon, specializing in gene-therapy vectors, is almost exclusively funded from the E85 million (US\$90 million) raised annually by the French Association for Neuromuscular Diseases.

Genopole has space for 60 start-ups and has already attracted 42. “We have passed the crucial threshold,” says director Pierre Tambourin. Unlike most other Parisian incubators, Genopole has a small (E1.2-million) seed fund to support companies in their early stages. “We have to compensate for the lack of ‘business angels’ in France,” says Tambourin.

SUBURBAN SPUR

The area south of Paris, including Evry, the home of Genopole, hosts one of the richest concentrations of public research in Europe. This includes many labs funded by the CNRS, France's national research agency, as well as the national research bases of the atomic energy commission (CEA) and the agronomic research institute (INRA), Paris-Sud University and some top-class engineering schools. Their proximity has attracted major companies such as telecoms giant Alcatel and the electronics group Thales, as well as many smaller firms.

The regional councils have been key players, funding not only Genopole, but also half of the E385-million budget for the third-generation Soleil synchrotron, currently under construction (see ‘Luring staff to Soleil’, opposite).

The region is now turning its attention to optics and photonics research. Alcatel set up Opticsvalley in 1999 to tap into the area's academic expertise. This non-profit association provides a technology-transfer service and runs a small incubator. Despite a downturn



in the telecoms industry, Opticsvalley has supported 12 start-ups, including a biophotonics project in collaboration with Genopole. “But we tend to refuse telecoms projects, however good they are, because capital riskers just aren't prepared to put money in,” says Jean-Claude Sirieys, who runs the incubator. “Instead we try to help restructure their business plan to serve the instrumentation market, for example.”

Plans are now under way to create a national optoelectronics platform based around semiconductor and biophotonics technology in the area, with a mix of public and private funding, employing 150 people.

Time will tell whether these efforts will be enough to attract investors away from other burgeoning European capitals. But a consistent long-term effort by central government to underpin the region's efforts is likely to be the key to a radiant future.

Sally Goodman is a freelance writer based in Paris.

Although the Soleil synchrotron is not yet in operation, recruitment is well under way.

Aventis alternatives

Aventis Pharma, France's premier pharmaceutical company, is planning to withdraw from one of its main research sites, at Romainville on the outskirts of Paris, as part of a management strategy to recentralize its research. The company will now focus its research effort in France on oncology and Alzheimer's disease, through its other Parisian centre at Vitry-sur-Seine.

Research on bone disease at Romainville has been spun out as a new biotech company,

ProSkelia, and the same solution is planned for part of the infectious-diseases research at Romainville, with the transfer of 100 posts. The site will become a technology park with space for biotech start-ups and larger companies. Although, according to Aventis Pharma, 400–500 jobs will be lost in the restructuring, they say that levels of employment will be maintained on the site in the long term through the science park.

But a group of Romainville employees

has put together an alternative proposal — christened ‘Nereis’ — which is arousing interest among potential partners. They envisage a public-private partnership undertaking research on drug candidates to combat infections, and on diseases currently neglected by the drug industry. Nereis would use Romainville's high-throughput screening facility to exploit molecules discovered in public laboratories, and to provide a drug-discovery training centre.

“The site has an extraordinary potential for drug discovery in terms of its technology platforms and expertise,” says Jean-Baptiste Léauté, director of industrial relations at Denis Diderot University in Paris, and adviser to the regional council. He believes that an alternative is possible for the site and is in discussions with ministries and public-research agencies to gauge support for the project.

S.G.

www.nereis-sante.com
www.aventispharma.fr

Pierre Tambourin: keen to support nascent companies.

